THE UNIVERSITY *of* EDINBURGH

Investigation of the tissue origins of cell-free DNA and candidate biomarkers for liquid biopsy

Danny Laurent

Submitted in fulfilment of the requirements for the degree of

PhD in Genomics & Experimental Medicine

The University of Edinburgh

2019

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

At the time of submission parts of this work have been submitted and currently being reviewed in a peer-reviewed journal BMC Medical Genomics:

Laurent D, Semple F, Lewis PJS, Rose E, Black HA, Forbes SJ, Arends MJ, Dear JW, Aitman TJ. Absolute measurement of the tissue origins of cell-free DNA in the healthy state and following paracetamol overdose. In review in BMC Medical Genomics.

Signed,

Danny Laurent

Abstract

Liquid biopsy is the sampling and analysis of analytes from various biological fluids to detect and diagnose cancer, for which it is an alternative to solid tissue biopsy for the detection of cancer. Compared to solid tissue biopsy, liquid biopsy is minimally invasive. Further, its ability to capture intra-tumoural and inter-metastatic heterogeneity in cancer makes it an attractive platform to complement solid tissue biopsy. Molecules of free circulating DNA in the bloodstream, termed cell-free DNA (cfDNA), can be used as an analyte for liquid biopsy. Various types of tumour-specific alterations can be detected in the cfDNA of cancer patients. To date, liquid biopsy assays based on the analysis of cfDNA are still lacking in consistency and precision. Clinical validity and utility have not been shown for the majority of assays. The composition of cfDNA in health and under particular physiological and pathological conditions still need to be better understood. Further improvements to diagnostic tools are required to enable detection of small numbers of molecules for liquid biopsy. Animal models may be useful as a platform to study cfDNA and cancer biology in liquid biopsy. The utilisation of an animal model may spare ethically sensitive clinical materials from patients and serve as an alternative when serial sampling from patients is challenging. For cancer biology, experiments in animal models are required to understand the mechanisms underlying the onset of malignancies and to identify methods to prevent, diagnose and treat diseases. In this PhD thesis, the tissue origins of cfDNA were investigated using tissue-specific reporter mouse models. Liver injury was induced in mice to investigate the effect of tissue injury on the composition of cfDNA. In addition,

the potential of the analyses of cfDNA as a biomarker of tissue injury in paracetamol (APAP) overdose patients was investigated. Absolute measurement of the tissue origins of cfDNA using ddPCR showed that myeloid, lymphoid and erythroid cells were major tissue contributors to cfDNA, adding up to ~78% of cfDNA pool in healthy tissue-specific reporter mice. A minor contribution was observed from hepatocytes (4.2%). No contribution was detected from muscle cells. Up to ~17.6% of contribution was unaccounted, indicating contribution of other tissues not studied in this project. Following APAP overdose in mice, total concentration of cfDNA increased by ~100-fold and the contribution of hepatocytes increased by ~20-fold. Similarly, a substantial increase in total cfDNA and liver-derived cfDNA was also observed in clinical samples from APAP overdose patients, highlighting the potential of the analyses of cfDNA as a biomarker of tissue injury in APAP overdose patients. In separate experiments, genomic analysis was performed on the cancer genome of the Pirc rat, a mutant rat of colorectal cancer (CRC), to screen for potential candidate biomarkers for liquid biopsy. Genomic analysis in the cancer genome of the Pirc rat showed candidate LOH biomarkers in chromosome 18, along with other types of somatic alterations in relevant cancer genes. These are potential markers of cancer development in future cfDNA studies in the Pirc rat. To conclude, successful analyses in rodent models showed the tissue origins of cfDNA in the healthy state and following APAP overdose, as well as potential candidate biomarkers of CRC for liquid biopsy.

Lay Summary

Liquid biopsy involves sampling and analysing substances from biological fluids to detect cancer. It is an alternative to invasive sampling from solid tissue for the detection of cancer, termed solid tissue biopsy. Compared to solid tissue biopsy, liquid biopsy is less invasive or harmful. Further, it is an attractive platform to complement solid tissue biopsy because it can capture variability within a tumour and variability within different tumours at different locations in the body. A DNA substance circulating freely in the bloodstream, called cell-free DNA (cfDNA), can be used for analysis in liquid biopsy. Many types of tumour-related DNA changes can be detected in cfDNA of cancer patients. To date, liquid biopsy tests based on the analysis of cfDNA are still lacking in consistency and precision. The extent of how well the tests can detect the presence or absence of diseases, and whether the tests provide clinically useful information, such as for diagnosis, treatment, management or prevention of diseases, have been incompletely demonstrated for the majority of liquid biopsy tests. The composition of cfDNA in healthy and other diseased conditions needs to be understood for optimal utility of the liquid biopsy approach. Animal models can be used to study cfDNA and the biology of cancer for liquid biopsy. The utilisation of animal model may spare ethically sensitive clinical materials from patients and serve as an alternative when repeated sampling from patients is challenging. For cancer biology, experiments in animal models are essential to understand the mechanisms underlying the beginning of cancer and to discover methods to prevent, diagnose and treat diseases. In this PhD thesis, the contribution of various

tissues that constitute cfDNA was investigated using healthy tissue-specific reporter mouse models. Tissue-specific reporter mouse models are mouse models with specific DNA alterations in tissues of interest. Liver injury was induced in mice to investigate the effect of tissue injury to the composition of cfDNA. Further, the potential of the analyses of cfDNA to detect tissue injury in paracetamol (APAP) overdose patients was investigated. Analysis of the contribution of different tissues to cfDNA in tissue-specific reporter mouse models using a state-of-the-art platform, called the ddPCR, showed that blood cells, which include myeloid, lymphoid, and erythroid cells, were major contributors to the composition of cfDNA, adding up to ~78% of the circulating cfDNA in the healthy state. Hepatocytes, a major component of the liver tissue, contributed around 4.2% to cfDNA in the healthy state. Muscle cell contributions were undetected. Up to ~17.6% of contribution was unaccounted, indicating likely contribution of other tissues not studied in this project. Following APAP overdose in mice, total concentration of cfDNA increased by ~100-fold and the contribution of hepatocytes increased by ~20-fold. This similar increase was observed in clinical samples from APAP overdose patients, highlighting the potential of the analyses of cfDNA as a biomarker of tissue injury in APAP overdose patients. Additionally, the analysis of whole DNA sequences, termed whole genome sequencing analysis, was performed in the tumour of large intestine from a genetically modified rat model of large intestinal cancer, called the Pirc rat. This analysis was performed to look for potential candidate changes in DNA to be used in liquid biopsy. Whole genome analysis of large intestinal tumours of the Pirc rat showed a large structural

alteration in DNA sequences, called the LOH in chromosome number 18, along with other types of tumour-specific changes in previously reported relevant cancer genes in human. This LOH alteration in cancer of the large intestine may serve as a potential alteration to be analysed in cfDNA for liquid biopsy. To conclude, this study successfully utilised mouse and rat models to investigate the tissue origins of cfDNA in the healthy state and following APAP overdose, as well as revealed potential biomarkers of colorectal cancer for liquid biopsy.

Acknowledgements

This PhD has been a long journey where I have learnt so many valuable things about science, working in an academic environment, and many other things in professional and personal life. I would like to acknowledge the people who have contributed along in my journey:

1. Prof. Tim Aitman. Thank you for the supervision all throughout my PhD, providing a workplace with so many expertise, and giving me an opportunity to learn about so many things, about science and more.
2. Dr. Fiona Semple. Thank you for the supervision on a day-to-day basis, and for helping me with even the simplest task when my brain was malfunctioning, and even more outside of the workplace.
3. Members of my thesis committee: Prof. Mark Arends, Prof. Charlie Gourley and Prof. Mary Porteous, for reviewing me, giving me a direction in research and helping me prioritise in getting a PhD. Prof. Mark Arends helped me a lot with histological analysis and interpretation of my samples.
4. Collaborators who helped shape my PhD into a viable project.
    a. Dr James Amos-Landgraf provided samples from Pirc rat.
    b. Prof. Stuart Forbes, Dr. James Dear and Dr. Philip Starkey Lewis helped me in the paracetamol tissue injury experiment and analysis in clinical samples.

c. Prof. Stuart Orkins provided EpoR Cre mice to investigate the contribution of erythroid cells to cfDNA

5. Members of the Aitman lab, past and present that I cannot mention them all. Dr. Holly Black taught me good laboratory work practices and was the PCR guru of the group. Elaine Rose helped me in processing hundreds of tissue samples and more. Dr. David Parry and Dr. Sophie Marion de Proce taught me how to do bioinformatics and whole genome sequencing analysis from the very beginning, and even witnessed how I succeeded in making my first “Hello World” Unix script. Dr. Neza Alfazema showed me how to isolate DNA from mouse tissues. I used the same method in 3 years and it was very efficient. Everyone else whose contribution I won’t list one by one but helped me throughout my study: David Ross, Jennifer Coe, Matthew Owens, Marjorie Barrier, Verdiana Martello, Dominique Balharry, Adam Jackson and Louise Pert.
6. People in the IGMM and CGEM:
    a. Dr. Alison Meynert helped answered so many questions about analysis of cancer genome sequencing
    b. Victoria Shaw helped me with my frustrations and even taught me to be brave to chase a dream.
    c. Susan Anderson, Helen Torrance, Joyce Begbie, Helen Caldwell.
7. The biomedical research facility (Gary Waugh, Scott Noble) and the Wellcome Trust Clinical Research Facility.

8. My sponsor the Indonesia Endowment Fund for Education for funding my PhD.

Finally, I am grateful to be involved and be able to learn from so many people mentioned and not mentioned in this acknowledgement, and now have an unforgettable experience from this PhD. Terima kasih.

Contents

Declaration……..ii

Abstract……iii

Lay Summary.....v

Acknowledgements viii

Contents……...xi

Table of Figures xv

Table of Tables.....xix

List of Abbreviations xxi

Chapter 1 Introduction.....1

1.1 Liquid biopsy: non-invasive detection of cancer 1

1.2 Definition and classifications of biomarkers 2

1.2.1 Types of DNA alterations used as biomarkers.....3

1.3 Cell-free DNA: an analyte in liquid biopsy 5

1.3.1 Applications of the analysis of cfDNA in other human diseases ...7

1.3.2 Current limitations of the analysis of cfDNA.....8

1.4 Development of technologies in liquid biopsy 9

1.4.1 Advances in pre-analytical methodologies of cfDNA analysis 9

1.4.2 Advances in analytical methodologies of cfDNA analysis.....11

1.4.3 Advances in post-analytical methodologies of cfDNA analysis...15

1.5 The role of rodent models in liquid biopsy 20

1.6 Thesis aims 22

Chapter 2 Materials and Methods.....24

2.1 Rodent Samples 24

2.1.1 Mouse breeding 24

2.1.2 Induction of APAP overdose in mice.....25

2.1.3 Rat Samples.....26

2.2 Human Samples 27

2.3 Sample pre-processing 28

2.3.1 Collection and processing of tissue samples 28

2.3.2 Collection and processing of blood samples 28

2.3.3 gDNA extraction 28

2.3.4 Cell-free DNA extraction 29

2.3.5 Bisulphite conversion of DNA and cfDNA 29

2.3.6 Quantitation of DNA 29

2.4 Histological analysis of tissue samples 29

2.4.1 Histological analysis of paraffin embedded tissue samples 29

2.4.2 Histological analysis of frozen tissue samples 30

2.5 Polymerase chain reaction 30

2.5.1 Quantitative polymerase chain reaction (qPCR) 30

2.5.2 Droplet digital polymerase chain reaction (ddPCR) 31

2.6 Fragment analysis of cfDNA 34

2.7 Analysis of protein biomarkers from plasma and serum samples 34

2.8 Absolute measurement of the tissue origins of cfDNA 34

2.9 Whole genome sequence analysis 35

2.9.1 Whole genome sequence generation and mapping to reference genome 35

2.9.2 Detection and validation of somatic LOH 36

2.9.3 Detection of somatic single nucleotide variants (sSNVs) and small insertion and deletion (indel) variants 37

2.9.4 Detection of structural and copy number variants 37

2.10 Statistics 38

Chapter 3 Absolute measurement of the tissue origins of cell-free DNA using healthy tissue-specific gene reporter mice 39

3.1 Introduction 39

3.2 Generation of conditional deletion in tissue-of-interest 41

3.3 Design and validation of assay to detect conditional deletion in genomic and cell-free DNA 42

3.4 Determination of DNA input requirement and minimum number of mice to analyse conditional deletion 45

3.5 Analysis of conditional deletion in gDNA from mouse tissues 46

3.6 Analysis of conditional deletion in cell-free DNA of healthy mice 50

3.7 Discussion 51

Chapter 4 The effect of tissue injury to the tissue origins of cell-free DNA 57

4.1 Introduction 57

4.1.1 Current understanding of the tissue origins of cell-free DNA following tissue injury 57

4.1.2 APAP overdose and tissue injury 57

4.2 Induction of APAP overdose in C57BL/6 mice 59

4.3 Analysis of Cre recombination in cfDNA of hepatocyte-specific reporter mice following APAP dosing 62

4.4 Analysis of cfDNA in APAP overdose patients 64

4.5 Discussion 66

Chapter 5 Screening for somatic mutations as candidate liquid biopsy biomarkers in the colonic adenocarcinoma of the Pirc rat 70

5.1 Introduction 70

5.1.1 Familial adenomatous polyposis 70

5.1.2 The Pirc Rat: a model for familial adenomatous polyposis 71

5.2 Histological validation of colonic tissue identity 72

5.3 Whole genome sequence generation and variant calling 73

5.4 Screening, detection and validation of somatic LOH variants in the Pirc rat cancer genome 81

5.5 Screening for somatic single nucleotide, small insertion and deletion variants in the Pirc rat cancer genome 93

5.6 Screening for structural and copy number variants in the Pirc rat cancer genome 95

5.7 Discussion 97

Chapter 6 General Discussion 103

6.1 Validation of ddPCR assay to quantify Cre-loxP recombination in tissue-specific reporter mice 103

6.2 Absolute measurement of the tissue origins of cfDNA in healthy mice 104

6.3 Measurement of the tissue origins of cfDNA following tissue injury 105

6.4 cfDNA analysis as a biomarker of tissue injury in APAP overdose patients 106

6.5 Genomic landscape of CRC in the Pirc rat 107

6.6 Follow up studies 108

References 111

Appendix ..... Error! Bookmark not defined.

A. Supplementary Figures ..... Error! Bookmark not defined.

B. Supplementary Tables 126

C. Statistical analysis 127

C.1 Comparison of body weight from different tissue-specific reporter mouse lines 127

C.2 Confirmation of Cre recombination in mouse tissues 128

C.3 Analysis of cfDNA and other liver biomarkers in clinical samples . 130

C.3.1 Comparison of total cfDNA in clinical samples 130

C.3.2 Comparison of liver-specific cfDNA in clinical samples 131

C.3.3 Associations of biomarkers in clinical samples 132

Table of Figures

Figure 1.2.1 Types of mutations.....4

Figure 1.3.1 The circulation of a cancer patient contains cfDNA from healthy and tumour cells.....7

Figure 1.4.2.1.4.1 Workflow of ddPCR.....13

Figure 1.4.2.2 Components and workflow of NaME-Pro.....15

Figure 1.4.3.1 A general workflow and representative sets of tools for the analysis of cancer genome sequencing.....17

Figure 1.5.1 Mechanism and breeding strategy of Cre-loxP recombination system.....21

Figure 1.6.1 Flowchart of the study design.....23

Figure 2.1.1.1 A schematic diagram of the mT/mG reporter gene before and after recombination.....25

Figure 2.8.1 A schematic representation of the overall strategy for absolute quantitation of the tissue origins of cfDNA.....35

Figure 3.2.1 Comparison of body weight from different tissue-specific reporter mouse lines at 10-12 week old.....42

Figure 3.3.1 Validation of ddPCR assays on gDNA containing 1lox or 2lox alleles.....44

Figure 3.4.1 Determination of ddPCR DNA input requirement.....46

Figure 3.5.1 Confirmation of Cre recombination in tissue-specific reporter mouse lines using ddPCR.....47

Figure 3.5.2 Visualisation of Cre recombination in liver and heart tissue sections of hepatocyte and cardiomyocyte-specific reporter mice.....48

Figure 3.5.3 Specificity of Cre recombination in tissue-specific reporter mouse lines across tissues. 49
Figure 3.6.1 The tissue origins of cfDNA in healthy mouse models. 51
Figure 4.1.2.1 APAP metabolism and mechanisms of hepatotoxicity. 58
Figure 4.2.1 The effect of APAP on biomarkers at different timepoints in C57BL/6 mice 61
Figure 4.3.1 The tissue origins of cfDNA following APAP overdose in mouse models 63
Figure 4.4.1 Specificity and validation of liver-specific methylation ddPCR assay. 64
Figure 4.4.2 Analyses of cfDNA and other liver biomarkers in clinical samples. 66
Figure 5.2.1 H&E-stained normal colon and colonic tumour tissues from the Pirc rat. 73
Figure 5.3.1 Pattern of heterozygosity in the Pirc rat genome from chromosome 1 to 4. 76
Figure 5.3.2 Pattern of heterozygosity in the Pirc rat genome from chromosome 5 to 8. 77
Figure 5.3.3 Pattern of heterozygosity in the Pirc rat genome from chromosome 9 to 12. 78
Figure 5.3.4 Pattern of heterozygosity in the Pirc rat genome from chromosome 13 to 16. 79
Figure 5.3.5 Pattern of heterozygosity in the Pirc rat genome from chromosome 17 to 20. 80

Figure 5.4.1 Screening for somatic LOH in a colonic tumour of the Pirc rat from chromosome 1 to 4. 82
Figure 5.4.2 Screening for somatic LOH in a colonic tumour of the Pirc rat from chromosome 5 to 8. 83
Figure 5.4.3 Screening for somatic LOH in a colonic tumour of the Pirc rat from chromosome 9 to 12. 84
Figure 5.4.4 Screening for somatic LOH in a colonic tumour of the Pirc rat from chromosome 13 to 16. 85
Figure 5.4.5 Screening for somatic LOH in a colonic tumour of the Pirc rat from chromosome 17 to 20. 86
Figure 5.4.6 Somatic LOH in chromosome 18 in two colonic tumours of the Pirc rat. 87
Figure 5.4.7 Validation of ddPCR assay to measure allele balance in the Pirc locus over a temperature gradient. 89
Figure 5.4.8 Screening of LOH across 46 colonic tumours from 21 ACI-Pirc rats and 2 ACI-WT rats. 91
Figure 5.5.1 Somatic single nucleotide variants found by three somatic variant callers from whole genome sequence of the tumours of Pirc rats 9 and 13. 94
Figure 5.5.2 Somatic small insertion and deletion variants found by three somatic variant callers from whole genome sequence of the tumours of Pirc rats 9 and 13. 94
Figure A.1 A representative plot of base quality of sequences across sequencing read. 123

Figure A.2 A representative plot of distribution of Phred quality score over all sequences. ……124
Figure A.3 A representative plot of GC content distribution from whole genome sequencing. ……...125

Table of Tables

Table 2.5.1.1.1 Oligonucleotides for amplification of 1lox and 2lox mT/mG gene. 31

Table 2.5.2.2.1 Oligonucleotides used in liver-specific methylation assay. .. 32

Table 5.3.1 Summary of whole genome sequence coverage and SNVs and small indels relative to the Brown Norway reference genome. 75

Table 5.4.1 Allelic and total sequencing depth at the Pirc locus 87

Table 5.4.2 Genomic coordinates of the somatic LOH regions in two Pirc rat colonic tumours. 88

Table 5.4.3 Screening of LOH across chromosome 18 of the Pirc rat cancer genome 92

Table 5.6.1 Structural and copy number variants of the Pirc rat cancer genome. 96

Table C.1.1 Median and standard deviation of the body weight of tissue-specific reporter mice. 128

Table C.1.2 p-values from comparison of mouse body weight between reporter lines with a Mann-Whitney U test. 128

Table C.2.1 Summary of percentage recombination and data distribution based on Shapiro-Wilk test in mouse tissues……129

Table C.2.2 p-value from a Mann-Whitney U test between tissues with and without recombination……...130

Table C.3.1.1 p-value from comparison of total cfDNA between clinical samples……..131

Table C.3.2.1 p-value from comparison of liver-specific cfDNA between clinical samples. 131
Table C.3.3.1 Spearman rho and p-value from analysis of associations between biomarkers in clinical samples……..132

List of Abbreviations

1lox	recombined floxed allele
1lox%	percentage of recombination of floxed allele
2lox	unrecombined floxed allele
acetyl-HMGB1	acetyl-high mobility group box 1 protein
ACI	August Copenhagen Irish
ACTB	human beta-actin
Actb	mouse beta-actin
ALB	albumin
Alb Cre	albumin Cre
AFLP	Amplification fragment length polymorphism
ALT	alanine aminotransferase
AOH	APAP overdose patient with high ALT
AON	APAP overdose patient with normal ALT
APAP	acetaminophen or paracetamol or acetyl-para-aminophenol
APAP-AD	APAP-protein adduct
APAP-GLU	non-toxic glucoronide metabolite of APAP
APAP-SUL	non-toxic sulfate metabolite of APAP
APC	human adenomatous polyposis coli
ARMS	amplification refractory mutation system
array-CGH	array-based comparative genomic hybridisation
AST	aspartate aminotransferase
BEAMing	beads emulsions amplification and magnetics
BRF	Biomedical Research Facility
BWA	Burrows-Wheeler Aligner
cfDNA	cell-free DNA
cfRNA	cell-free RNA
CNV	copy number variant
CRC	colorectal cancer
CSF	cerebrospinal fluid
CSF-1	colony stimulating factor 1
CTC	circulating tumour cells
ctDNA	circulating tumour DNA
cTnT Cre	cardiac troponin T Cre
ddPCR	droplet-digital PCR
ddPCR	droplet-digital PCR
ELISA	Enzyme-linked immunosorbent assay
ENU	N-ethyl-nitrosourea

EpoR Cre	Erythropoietin receptor Cre
FAP	familial adenomatous polyposis
gDNA	genomic DNA
gDNA	genomic DNA
GLDH	glutamate dehydrogenase
GSH	glutathione
GST	glutathione-s-transferase
H&E	hematoxylin and eosinophil
hCD2-iCre	human CD2 promoter-iCre
HV	healthy volunteer
IP	intraperitoneal
LOH	loss-of-heterozygosity
LysM Cre	lysozyme 2 Cre
MCK Cre	muscle creatinine kinase Cre
mG	cell-membrane localised enhanced green fluorescent protein
MS-NAME	methylation-specific nuclease-assisted minor allele enrichment
mT	cell-membrane localised tdTomato
NaME-PrO	nuclease-assisted minor-allele enrichment using overlapping probe
NAPQI	N-acetyl-p-benzoquinone imine
NIPT	non-invasive prenatal testing
RFLP	Restriction fragment length polymorphism
SNP	single nucleotide polymorphism
SNV	single nucleotide variant
sSNV	somatic single nucleotide variant
sSNV	somatic single nucleotide variant
SULT	sulfotransferase
SV	structural variant
Tam-Seq	tagged-amplicon deep sequencing
UGT	UDP-glucoronosyltransferase
UoE	University of Edinburgh
WES	whole exome sequencing
WGS	whole genome sequencing
WHO	World Health Organisation

Chapter 1 Introduction

1.1 Liquid biopsy: non-invasive detection of cancer

Liquid biopsy is the sampling and analysis of analytes from various biological fluids to detect and diagnose cancer, a disorder characterised by abnormal and uncontrolled cell growth. It serves as an alternative to solid tissue biopsy, an invasive extraction of solid tissue to detect cancer that is risky, painful and, in some cases, unavailable [1, 2]. Analytes in liquid biopsy include circulating tumour cells (CTCs), circulating cell-free DNA (cfDNA) and RNA (cfRNA), extracellular vesicles, proteins and other metabolites [1]. These analytes are derived predominantly from blood but also other accessible fluids, such as cerebrospinal fluid (CSF), urine, ascites fluid [1, 3].

The gold standard method for genetic profiling of cancer is tissue biopsy [4]. It provides morphological information of the tumour and has high diagnostic certainty [3]. However, due to its invasive nature, tissue biopsy imposes higher risk to patients compared to liquid biopsy. Moreover, tissue biopsy may fail to capture the heterogeneity of cancer within a tumour tissue and between tumours at different locations, thus affecting the accuracy of the test [4]. Together, tissue and liquid biopsy complement each other and provide the highest diagnostic standard for cancer patients [3].

More studies are required to show the potential of cancer liquid biopsy and overcome its limitations [3]. The clinical applications of liquid biopsy potentially include detection and localisation of neoplastic growth, prediction of tumour stage and prognosis, genotyping of tumour and identification of potential

targeted therapeutics, monitoring treatment efficacy and resistance, as well as identification of early-stage recurrence [5]. Despite these potential applications, liquid biopsy assays are currently still lacking in consistency and precision. Clinical validity and utility have not been shown for the majority of assays. The composition of cfDNA analytes in health and under particular physiological and pathological conditions need to be better understood. Finally, further improvements to diagnostic tools are required to enable detection of small amounts of molecules for liquid biopsy [1].

1.2 Definition and classifications of biomarkers

Understanding different types of biomarkers is important to discuss liquid biopsy approaches in details [2]. A biomarker is an objectively measured characteristic that describes a normal or abnormal state in an organism by analysing biomolecules such as DNA, RNA, protein, peptide and chemical modifications of biomolecules. The World Health Organisation (WHO) defined a biomarker as any substance, structure or process that can be measured in the body or its product and influence or predict the incidence or outcome of diseases. A cancer biomarker may measure the risk of developing cancer in a specific tissue or, alternatively, may measure the risk of cancer progression or potential response to therapy. Biomarkers play an important role in the clinical management of cancer patients [6].

Cancer biomarkers can be classified into three categories based on their utility. Predictive biomarkers predict response to specific therapeutic interventions. Prognostic biomarkers may not be linked to therapeutic decisions but aim to

inform clinicians about the risk of clinical outcomes such as cancer recurrence or disease progressions in the future. Diagnostic biomarkers identify whether a patient has a specific disease condition [6].

1.2.1 Types of DNA alterations used as biomarkers

Different types of alterations in DNA may be used as a biomarker of cancer. This may include alterations in nucleotide sequence of DNA, termed mutations, or other modifications in DNA, including CpG dinucleotide methylation changes [7, 8].

A mutation can be distinguished into two categories based on its location in the body: germline and somatic mutations. A germline mutation occurs in the germ line, tissue that forms reproductive cells during the course of development. A germline mutation may be passed onto some or all individuals of the next generation. A somatic mutation occurs in a single cell or a population of cells in developing somatic tissue and cannot be transmitted onto individuals of the next generation [9].

Based on the types of alterations on DNA sequences, a mutation can be classified into point mutations and small deletions, as well as chromosomal abnormalities (**Figure 1.2.1**). A missense mutation involves a change in a single nucleotide (also called single nucleotide variants, SNV). A nonsense mutation is a change in nucleotide bases that changes the amino acid into stop signal in protein-coding genomic DNA (gDNA) sequence. A frameshift mutation can be caused by additions or deletions of nucleotides that changes the reading frame in which DNA is translated. Chromosomal abnormalities

include large structural changes, such as inversion, duplication, deletion, translocation and insertion [10].

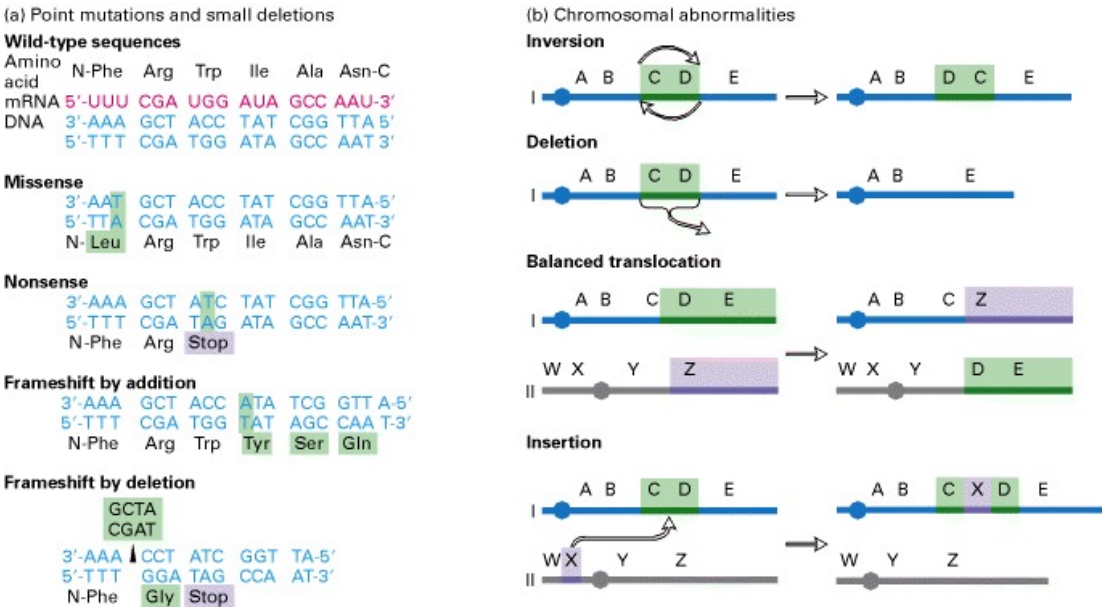

Figure 1.2.1 Types of mutations.
The types of mutation can be classified into point and small deletions, as well as chromosomal abnormalities. (a) point mutation and small deletions typically affect the function of only one gene. (b) Chromosomal abnormalities affect large segments of DNA. Illustration taken from [10].

DNA methylation is a modification of DNA base by the addition of a methyl group by DNA methyltransferase enzymes. This modification predominantly occurs at CpG dinucleotides sequences in mammalian cells. The levels of DNA methylation at individual CpG site as well as the distribution of methylated and unmethylated CpGs are linked to development state and cellular differentiation [11], leading to specific profiles of methylation pattern in different tissues of mammalian organisms [12]. Perturbation of CpG dinucleotide methylation may lead to changes in gene expression or genomic instability and lead to cancer [11].

Analysis of CpG dinucleotide methylation may be performed following a bisulphite conversion procedure. This procedure converts unmethylated cytosine bases into uracils, while leaving methylated cytosines as cytosines. Changes in this DNA sequence can then be detected by a variety of subsequent analyses, such as sequencing. The bisulphite conversion procedure may degrade a significant amount of DNA, up to 95%. Therefore, multiple factors, including DNA input amount, are controlled to optimise conversion and reduce loss of DNA [8].

For the detection of cancer, analyses of different types of tumour-specific alterations come with different advantages. Analyses of somatic mutations in liquid biopsy may offer high specificity because these alterations are present only in genomes of cancer or pre-cancerous cells and not in DNA of normal cells of the same individuals [13]. On the other hand, analyses of epigenetic alterations and nucleosomal mappings in cfDNA may show the presence of tumour and locate its tissue of origin [1, 14, 15].

1.3 Cell-free DNA: an analyte in liquid biopsy

The presence of DNA circulating in human plasma was first described in 1948 [16]. These molecules are termed cell-free DNA (cfDNA). Aside from the plasma, cfDNA can also be isolated from other biological fluids such as urine, CSF, and ascites fluid [17]. In the plasma, a proportion of cfDNA circulates as nucleosome or chromatosomes, nucleosome that are bound to histone H1 protein, and another as free DNA [14, 17]. The size of cfDNA from healthy cells is postulated to be between 160 and 200bp [18, 19] with different estimates between methodologies including XXX and YYY. For example, using high

throughput sequencing, the size of cfDNA is on average of 167bp, originating from approximately 147bp from nucleosome and 20bp linker histone [14]. cfDNA that originates from tumour in cancer patients and foetus in the circulation of pregnant women is on average shorter than cfDNA from healthy cells [18, 20, 21].

cfDNA is thought to be released by cells into biological fluids by a mixture of cellular breakdown mechanisms, such as apoptosis and necrosis, as well as active DNA release mechanism from living cells [17, 19, 22]. Studies suggest that apoptosis is the main mechanisms of cfDNA release in both healthy and diseased states [19]. This is supported by the fact that cfDNA fragments consisted of a ladder pattern of 180-200bp or multiples thereof when analysed using DNA electrophoresis and sequencing [18, 19].

Analysis of the tissue origins of cfDNA in human has indicated that cfDNA originates predominantly from apoptosis of hematopoietic cells [14, 23, 24], with minimal contribution from other tissues, such as liver [25] and endothelial cells [26]. The circulation of cancer patients may contain cfDNA that derives from tumours, termed circulating tumour DNA (ctDNA) (**Figure 1.3.1**). The proportion of ctDNA in the circulation of cancer patients varies depending on types of tumour and tumour burden [1, 22]. The circulation of pregnant women contains both cfDNA from the mother and the foetus [27].

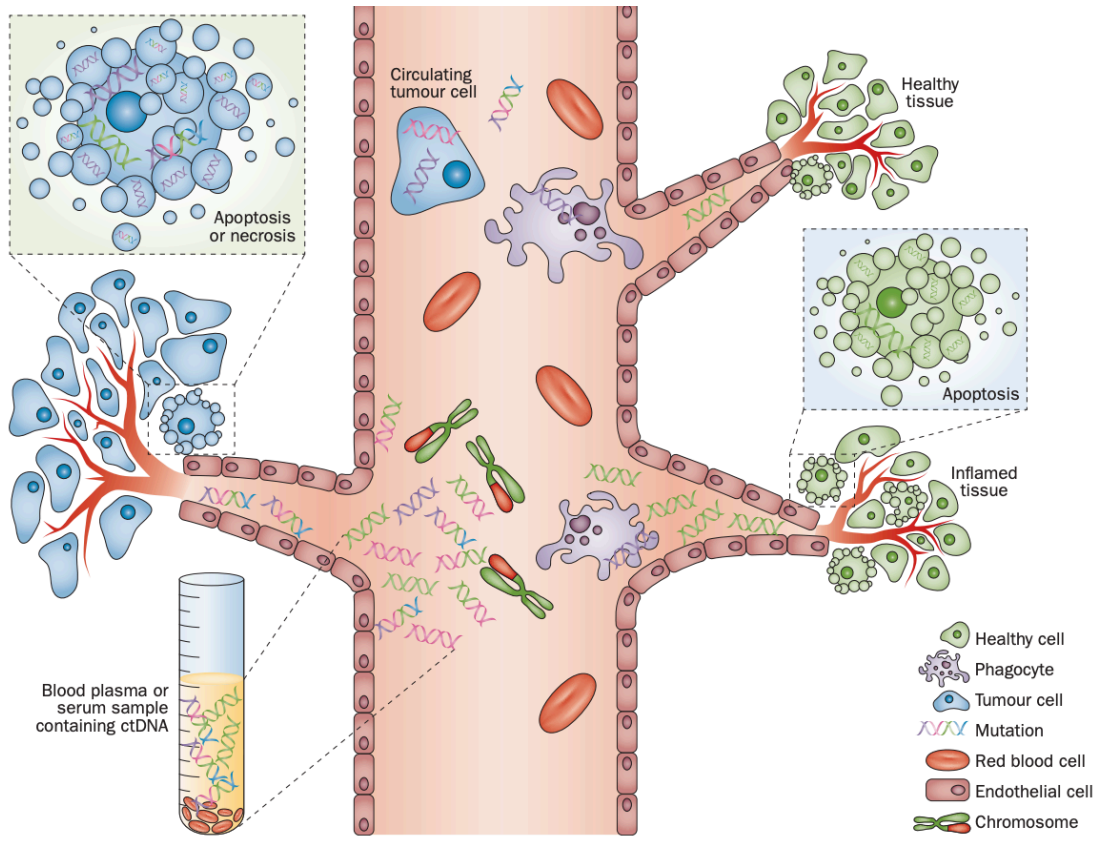

Figure 1.3.1 The circulation of a cancer patient contains cfDNA from healthy and tumour cells.
In the circulation of a cancer patients, DNA is released by healthy cells (shown in blue) and tumour cells (shown in green), mainly through cellular death mechanisms. DNA from tumour cells, ctDNA may contain tumour-associated alterations, including point mutations and structural variations. Illustration taken from [22].

1.3.1 Applications of the analysis of cfDNA in other human diseases

The presence of tumour DNA in the circulation of cancer patients [28], and foetal DNA in the circulation of pregnant women [27] formed the basis for the two main applications of the analysis of cfDNA: in the detection of cancer and foetal abnormalities. The detection of foetal abnormalities using the analysis of cfDNA is termed non-invasive prenatal testing (NIPT). NIPT is an alternative to invasive prenatal testing that includes an invasive amniotic fluid sampling that poses a risk of miscarriage [29]. NIPT is now routinely used in the clinic

for detection of foetal aneuploidy, mainly for the detection of trisomy 13, 18 and 21 [29, 30]. Proof-of-concepts experiments for further development of NIPT are currently ongoing. This include the analysis of whole foetal genome, methylome and transcriptome [29].

Other than cancer liquid biopsy and NIPT, there are other potential applications of the analysis of cfDNA, such as tissue health in solid organ transplantation [31] and other internal tissue injury [32, 33]. For monitoring on the health in solid organ transplantation, the analysis of cfDNA is an alternative to donor organ biopsy. Donor organ biopsy is typically performed in response to suspicion of graft pathology or as a part of a routine monitoring protocol. Proof-of-concepts experiments for monitoring on tissue health in solid organ transplantation have been shown for various tissues, including kidney, liver, heart, pancreas and lung [31]. The applications of the analysis of cfDNA for other internal tissue injury have been shown for injury in brain following traumatic brain injury, pancreas following pancreatitis, and heart following myocardial infarction [32, 33].

1.3.2 Current limitations of the analysis of cfDNA

Currently, both physiological and technical factors are limiting the implementation of cancer liquid biopsy based on the analysis of cfDNA in the clinic. The composition of cfDNA in healthy individuals and how it changes during physiological processes are not well understood [1]. Knowledge of biological factors and mechanisms that influence the size distribution of cfDNA are still limited [20]. Plasma often contains only low levels of cfDNA. This limits

the number of input molecules, types of tumour-specific alterations to be analysed and, particularly, affects the detection of mutant DNA fragments with low of variant allele fractions. Finally, the implementation of cfDNA is still limited by technical limitation of analytical limit of detection of different methods being used [1].

1.4 Development of technologies in liquid biopsy

A liquid biopsy test consists of three distinct phases, including the pre-analytical, analytical and post-analytical phase. The pre-analytical phase of liquid biopsy testing includes the collection, transport and pre-processing of samples. The analytical phase comprises of the laboratory testing of samples with the selected methods. The post-analytical phase consists of data analysis, interpretation and reporting. A standardised and optimised analytical workflow is important to allow for robust and reproducible cfDNA analysis [34]. This section describes advances made in these three phases of testing in cancer liquid biopsy, specifically using cfDNA.

1.4.1 Advances in pre-analytical methodologies of cfDNA analysis

Variations in pre-analytical processing of samples significantly affect results of cfDNA analysis. Optimal and standardised sample pre-processing methods are important to minimise lysis of peripheral blood cells prior to cfDNA analysis [35]. Pre-analytical considerations of cfDNA samples include sample sources of cfDNA, blood collection and preservation, cfDNA isolation and storage [36]. cfDNA can be obtained mainly from either plasma or serum samples.

The choice between serum and plasma as the optimal sampling specimen is controversial [7]. Measurement of total cfDNA showed that serum samples contained more cfDNA than plasma samples. Further, estimation of total amounts of telomeres shows significantly higher amount in cfDNA from serum compared to plasma [37]. Analysis of DNA fragment size showed that cfDNA from serum samples has larger numbers of longer fragments of cfDNA than from plasma samples. This suggested the occurrence of lysis of leukocytes in cfDNA from serum samples during sample processing. However, serum samples are still a useful source of cfDNA. A study showed that there is no significant difference in terms of sensitivity and specificity of cfDNA analysis from serum or plasma samples for cancer detection [38].

The concentrations of extracted cfDNA are affected by different technical factors, such as anticoagulants, blood collection tubes and the time of delay between blood draw and centrifugation. The total concentration of cfDNA increase overtime prior to centrifugation in EDTA-stabilised blood. Specialised blood collection tubes, such as cell-free DNA BCT (Streck) and PAXgene blood DNA tube (Qiagen) contain anticoagulants that prevent lysis of blood cells for several days at room temperature [38]. It is recommended to use specialised blood collection tubes, or in case of EDTA tubes, process cfDNA within 2 to 6 hours of blood collection to prevent the release of gDNA from leukocytes [36].

The majority of studies utilise a single centrifugation or double centrifugation steps to separate plasma or serum from blood samples. The majority of these studies showed that there is no significant difference in between a single and

double centrifugation steps to DNA yield [38]. However, a single study suggested that a double centrifugation protocol may be better in removing contaminating nuclear DNA than single centrifugation step [39]. A double centrifugation protocol is recommended to allow for complete separation of cellular fractions and avoid mechanical lysis of leukocytes [36].

There is an effect of storage temperature to the concentrations of extracted cfDNA in EDTA-stabilised blood when processed within 6 hour [38]. The concentrations of cfDNA are not affected for 2 weeks when plasma samples are stored in -80°C. But a significant decrease is shown when plasma samples are stored in more than 37°C [40]. It is recommended to freeze plasma samples at -20 or -80°C, not store plasma samples at 2-8°C for longer than 24 hour, and avoid frequent freeze thaw cycles to prevent increased fragmentation and reduced cfDNA yield [36].

Methods of cfDNA extraction vary between in house protocols, such as phenol chloroform extraction [41] and different commercially available kits [38]. The majority of cfDNA extraction protocols are either column- or magnet-based. The most commonly used commercial kit is the QIAamp circulating nucleic acid (Qiagen). This kit has been reported to produce high cfDNA yield. An automated extraction protocol is advantageous to minimise hands-on time and aid in standardisation [42].

1.4.2 Advances in analytical methodologies of cfDNA analysis

Early advances in cfDNA analyses for liquid biopsy was focused on simple assessment of cfDNA concentration in the circulation [43]. The low concentration of cfDNA extracted from plasma makes quantification using

spectroscopy and fluorometry problematic. Quantitative PCR (qPCR) assays are more frequently used for measurements of total cfDNA [44]. Total concentration of cfDNA can be quantified using a single-locus qPCR [45], or qPCR of repeat DNA sequences, such as qPCR of ALU sequences [46]. The size of cfDNA fragments can be analysed using DNA fragment analyser, qPCR, or more accurately using next-generation sequencing [20, 47]. These simple methodologies for the assessment of cfDNA have a limited utility for detection of cancer. This in turn leads to the development of specific approaches for the analysis of cfDNA for cancer detection [43].

There are two main approaches in analysis of cfDNA for detection of cancer. A targeted approach includes the analysis of known tumour-specific alterations. The alteration may be a frequently occurring driver mutation with implications for therapeutic decision. An untargeted approach does not require knowledge of any specific alteration associated with tumours. An untargeted approach is often used to discover novel tumour-specific alterations and useful for detection of molecular resistance as well as identification of new therapeutic targets [43].

Development of methodologies for targeted analysis of cfDNA has allowed for identification of mutant alleles at very low frequencies, down to 0.02%. These methodologies include amplification refractory mutation system (ARMS), beads emulsions amplification and magnetics (BEAMing), droplet digital PCR (ddPCR), as well as personalised panel for cancer profiling, such as cancer personalised profiling by deep sequencing (CAPP-Seq) and tagged-amplicon deep sequencing (TAm-Seq) [43, 48, 49]. ARMS is a simple method for

detecting mutations involving small base changes or deletions. It utilised sequence-specific PCR primers that only amplifies DNA when the target allele is present [50]. BEAMing and ddPCR utilise magnetic and droplet emulsion partitioning, respectively, to perform single molecule amplification, with ddCPR (**Figure 1.4.2.1**) being the major method to detect genomic alterations in cfDNA [51, 52]. Analyses of cfDNA with personalised panel combines a multiphase bioinformatics approach to identify targets, an optimised library preparation methods, and sequencing at a deep coverage [43].

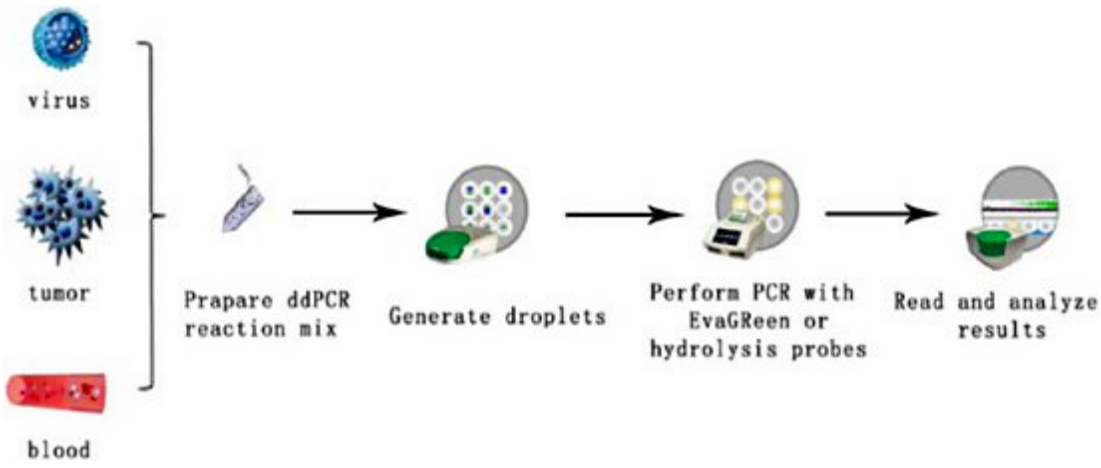

Figure 1.4.2.1.4.1 Workflow of ddPCR.
DNA template is extracted from virus, tumour or blood. Following extraction, DNA template is mixed with PCR reaction mix and partitioned randomly into individual droplets of approximately 20000 droplets. Amplification reaction is performed with double-stranded DNA dye EvaGreen or hydrolysis probe. Fluorescent signal from each individual droplet is read after amplification, giving positive or negative signal. Illustration taken from [52].

Unlike targeted approach, genome-wide analyses of liquid biopsy are applicable to all patients because these analyses do not rely on recurrent genetic alterations. Further, genome-wide analyses are able to detect not only small alterations, but also tumour-associated copy number alterations. Examples of methodologies in this approach include array-based comparative

genomic hybridisation (array-CGH), whole exome sequencing (WES) and whole genome sequencing (WGS). These methodologies are less cost-effective than targeted methodologies. However, with the reduction of sequencing costs and improvement of turnaround time, as well as emergence of novel sequencing technologies with improved accuracy, this approach will eventually be widely available for clinical purposes [43].

The sensitivity of both targeted and whole-genome approach for the detection of cancer is affected by low abundance of ctDNA within an excess of background cfDNA originating from healthy cells. To overcome this challenge, methodologies have been developed to allow for enrichment of ctDNA by removing excess background cfDNA from healthy cells, such as nuclease-assisted minor-allele enrichment using overlapping probe (NaME-PrO) (**Figure 1.4.2.2**) and methylation-specific nuclease-assisted minor allele enrichment (MS-NaME), for detection of somatic mutation and tumour-associated CpG dinucleotide methylation changes respectively. Both technologies utilise probes that specifically bind to background cfDNA, but not fully bind to ctDNA due to mismatches from somatic mutations or methylation changes. A duplex-specific nuclease digests dsDNA, resulting in enrichment of alleles with tumour-specific alterations [53, 54].

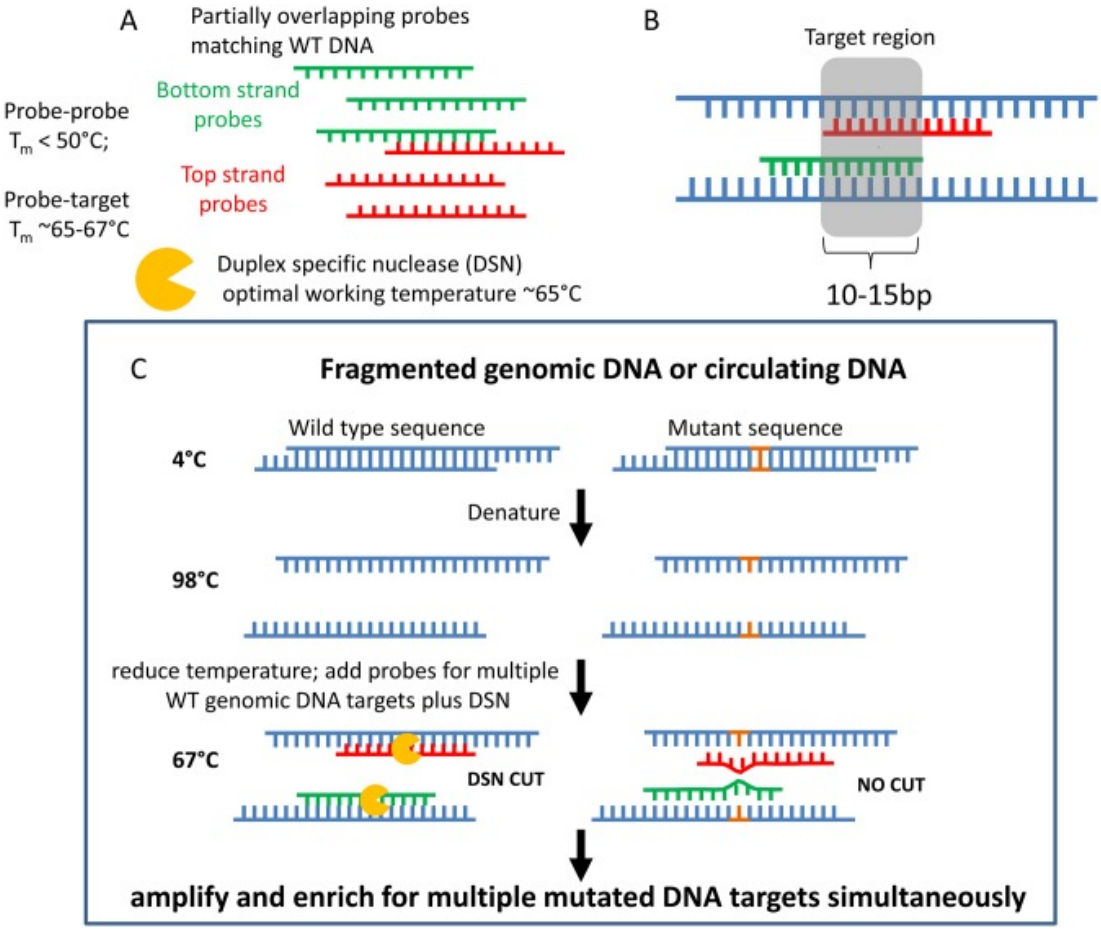

Figure 1.4.2.2 Components and workflow of NaME-Pro.
(A) Components of NaME-Pro consists of specific probes that bind to bottom and top strands of wildtype DNA, and a duplex-specific nuclease. (B) Probes hybridise to target region of 10-15bp in length. This overlapping region defines the target DNA region to be enriched. (C) The workflow of NaME-Pro consisted of denaturation and incubation of DNA with NaME-Pro components at ~67°C. Illustration taken from [53].

1.4.3 Advances in post-analytical methodologies of cfDNA analysis

The advancement of technologies and bioinformatics algorithm for the analysis of cfDNA in liquid biopsy is rapid. This is supported by the maturation of massively parallel sequencing technologies that allows for the characterisation of cfDNA at a single base resolution and genome-wide scale [55]. Best practices workflow for algorithms of data pre-processing and discovery of

somatic short variants, such as single nucleotide variants and small indels, are now available [56–58]. Recent advances in bioinformatics also allow for estimation of ctDNA fraction in cfDNA, non-invasive detection of copy number aberrations, genomic rearrangements and methylation analysis for the detection of cancer [55].

A number of tools for analysis of variant discovery from the cancer genome is now widely available (**Figure 1.4.3.1**) [59]. Raw sequence data is typically aligned to reference genome using alignment tools, such as Burrows-Wheeler Aligner (BWA) [58], to produce alignment files (BAM files). Several algorithms are then available to call somatic variants from alignment files. Most of these tools compare variant allele fraction in the cancer genome with paired normal genome. Different tools are utilised to discover different types of somatic variants, such as single nucleotide variants (SNVs), short indels variants, structural variants (SVs) and copy number variants (CNVs). These variant callers produce vcf files containing a list of discovered variants. Variants can then be filtered and annotated for further interpretation [59].

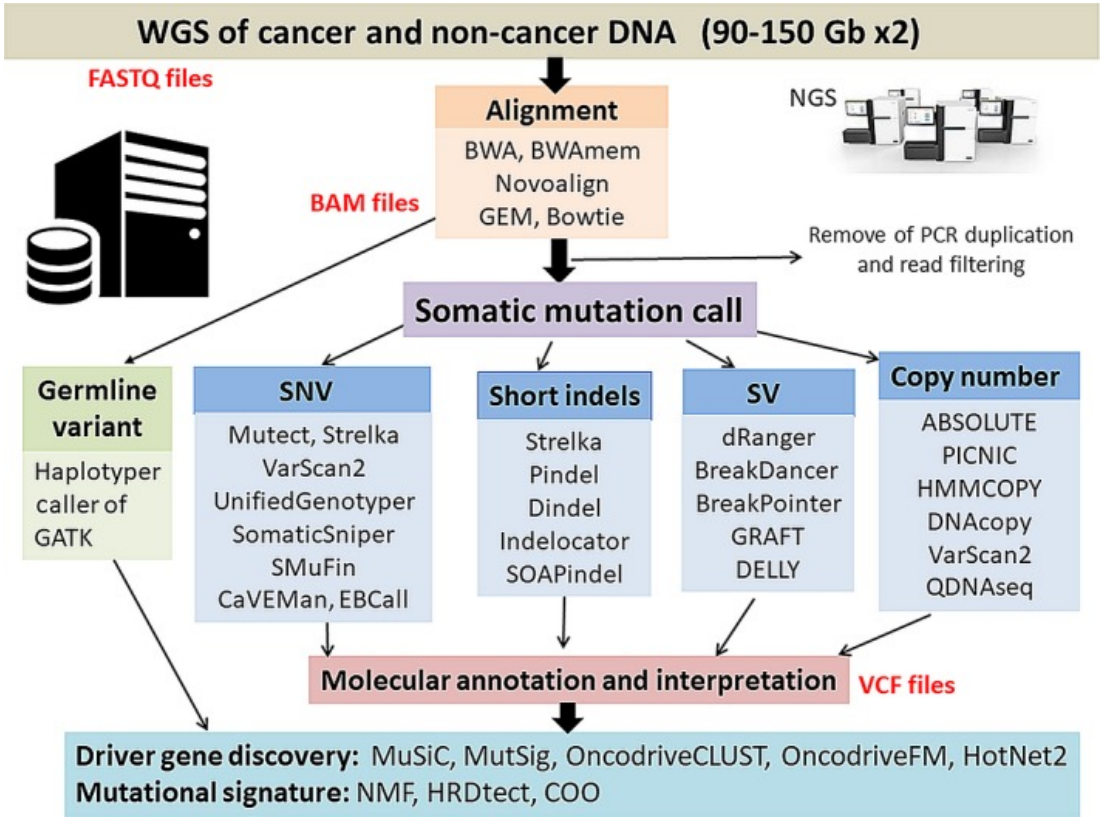

Figure 1.4.3.1 A general workflow and representative sets of tools for the analysis of cancer genome sequencing.
Raw sequence data is aligned to generate BAM alignment files. PCR duplicates are removed from BAM files prior to different types of somatic variant calling, such as SNV, short indels, SV and copy number analysis. Variants are annotated and interpreted. Illustration taken from [59].

Accurate estimation of ctDNA within the background of cfDNA from healthy cells has proved challenging. The fraction of ctDNA in cfDNA may be estimated by conventional PCR amplification of multiple oncogenes. However, the heterogeneity of tumour cells complicates PCR based estimation of ctDNA fraction. For example, some variants are only present in subclones of a tumour, so that lower fractions of these variants are present in the circulation, leading to inaccurate estimation of ctDNA fraction. Estimation of ctDNA fraction can be analysed using two bioinformatics approaches in massively parallel sequencing data: the mutant allele approach and the loss-of-heterozygosity (LOH) approach. The mutant allele approach utilises a predetermined set of

mutations commonly found in certain cancers or found in matching tumour samples. The mutant allele fraction can be expressed as the number of sequencing reads containing mutant reads over total alleles. The LOH approach takes the advantage of the presence of LOH in tumours. Sequencing reads carrying heterozygous single nucleotide variants reflect non-tumour fraction of cfDNA, whereas reads carrying LOH variants reflect ctDNA [55].

Non-invasive detection of copy number variants in cfDNA may serve as a promising cancer detection biomarker because copy number aberrations can be found in almost all tumours. There are two main approaches for detection of copy number variants in sequence data. The first approach divides the genome into equally sized bins, and estimates copy number gains or losses by the increase or decrease of genomic representation in cfDNA. The other approach examines aberrant changes in chromosomal arms. The bin approach is more generalised and can be applied to all cancer, whereas the chromosomal arms approach can only be applied to cancer types that primarily affect chromosomal arms, such as breast and colon cancer [55].

As with copy number aberrations, genomic rearrangements, such as BCR-ABL fusion in chronic myeloid leukaemia, is common in many cancer types. The detection of gene fusion in cfDNA utilises mate-pair library sequencing to identify reads containing genomic rearrangement breakpoints. These breakpoints can be validated by designing PCR assays across the breakpoints. Detection of genomic rearrangements has been shown to be specific for cancer patients, however, it requires high sequencing depth [55].

Finally, advances have been made in the analysis of DNA methylation in cfDNA. CpG dinucleotide methylation can be analysed in cfDNA by treating DNA with sodium bisulphite. This converts unmethylated cytosines into uracils while leaving methylated cytosines unchanged. PCR amplification can be performed using primer-pairs specific for bisulphite-converted DNA. Amplification of bisulphite-converted DNA will convert uracils into thymines, whereas methylated cytosines will remain as cytosines. Sequencing of bisulphite converted DNA and comparison of its results to original DNA sequence would reveal the methylation status. During bisulphite sequencing, unmethylated cytosines are sequenced as thymines, whereas methylated cytosines are read as cytosines. The presence of cytosines indicates the presence of methylation in the original DNA sample. Multiple bisulphite sequencing alignment tools, such as Bismark and BSmooth, are now available [55, 60–62].

Aside from DNA bisulphite conversion, amplification and sequencing, methylation status can be shown by other methods. Profiling of methylation status across whole genome can be performed by chromatography. Separation of methylated cytosines from other components of hydrolysed DNA and measurement of its fraction can show global levels of methylation across the whole genome. Global methylation status can also be shown using enzyme-linked immunosorbent assay (ELISA), using antibody specific for methylated cytosines. Detection of the methylation status of specific fragments of DNA may also be performed using methylation-specific fingerprinting methods, such as amplification fragment length polymorphism (AFLP) and

restriction fragment length polymorphism (RFLP). Detection of methylation status across the genome can be performed using hybridisation array as an alternative to bisulphite sequencing, or bisulphite sequencing following enrichment using methyl-binding proteins or antibody immunoprecipitation [63].

1.5 The role of rodent models in liquid biopsy

Genetically modified animals are organisms in which specific genes have been altered to create models for human and animal diseases. A transgenic model is defined as an animal that carried one or more foreign gene, deliberately through insertion into the animal's genome. The foreign gene is constructed using recombinant DNA technology. In cancer, control of oncogenes and tumour suppressor genes can be controlled spatial and temporally in animal models using tissue-specific and inducible expression of Cre-loxP recombination system (**Figure 1.5.1**), a bacteriophage P1 site-specific recombination system that enables excision of DNA fragments between two loxP sites [64–66].

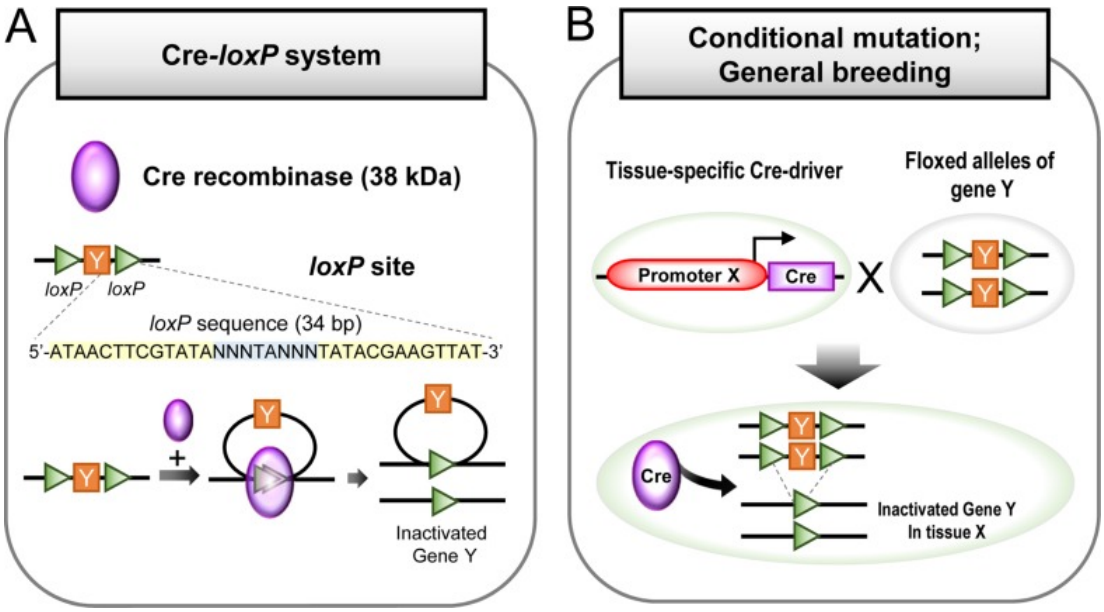

Figure 1.5.1 Mechanism and breeding strategy of Cre-loxP recombination system. (A) Mechanism of Cre-loxP system. Cre recombinase (38kDa) recognises loxP sequences and facilitates excision of gene Y, that is flanked in between the two loxP sites, leaving one loxP site following recombination. (B) Mouse carrying Cre recombinase allele under a specific promoter X is bred with mouse carrying target alleles between two loxP sites (floxed alleles). The litter (F1) may inherit both Cre recombinase gene and target alleles, resulting in inactivation of target gene Y in a specific tissue X. Illustration taken from [66].

The mouse is the most common model organism for preclinical studies. Mouse models have many advantages over other model organism. A comparison between mouse and human genome showed that mouse shares a majority of its protein-coding genes with human [67], with 40% similarity at DNA sequence level to the human genome [68]. Genetic and molecular tools are available for modification of the mouse genome. The small animal size facilitates study for larger numbers making it a cost-efficient model [65].

Similarly to mouse models, resources for the genetic manipulation of rat models are now available [69]. Rat models offer the advantage of easier monitoring of physiology and ability to perform of surgical procedures. Additionally, rats are more intelligent than mice and capable of learning a wider variety of tasks important to cognitive research. The choice between mouse or

rat model for studying a particular condition would be specifically depend on which model is more biologically appropriate [70].

There are advantages of animal models as alternative platform to study cfDNA and cancer biology in liquid biopsy. Animal models can be used to optimised liquid biopsy approaches and spare ethically sensitive clinical materials from patients [71]. Studies in animal models can serve as an alternative when serial sampling from patients is challenging [72]. For cancer biology, experiments in animal model is essential to understand fundamental mechanisms underlying the onset of malignancies and to discover methods to prevent, diagnose and treat diseases [65].

1.6 Thesis aims

In this PhD project, genomic analyses were performed on rodent models to study the tissue of origins of cfDNA and the genomic landscape of colorectal cancer (CRC) (**Figure 1.6.1**).

- Chapter 3 aimed to establish and utilise tissue-specific reporter mouse models in investigating the tissue origins of cfDNA in the healthy state. In this chapter, assays for absolute quantification of the tissue origins of cfDNA were established. Further, the optimum input amount of cfDNA for analysis of tissue origins is investigated. Finally, the absolute measurement of the tissue origins of cfDNA was shown for six cell types.
- Chapter 4 aimed to investigate the effect of tissue injury to the tissue origins of cfDNA in tissue-specific reporter mice. Paracetamol-induced liver injury was performed in mice to confirm expected increase in total

cfDNA and contribution of liver to cfDNA. Further, comparison of cfDNA analysis and existing protein biomarker was performed to show the utility of cfDNA analysis as a biomarker of paracetamol-induced liver injury in clinical samples.

- Chapter 5 aimed to screen for potential candidates of biomarkers for detection of CRC using WGS of a mutant rat model. Various types of tumour-specific alterations, including single nucleotide variants, short indels, structural and copy number variants, as well as LOH variants were screened in CRC of the Pirc rat. Additionally, screened LOH variants were validated using ddPCR assays.

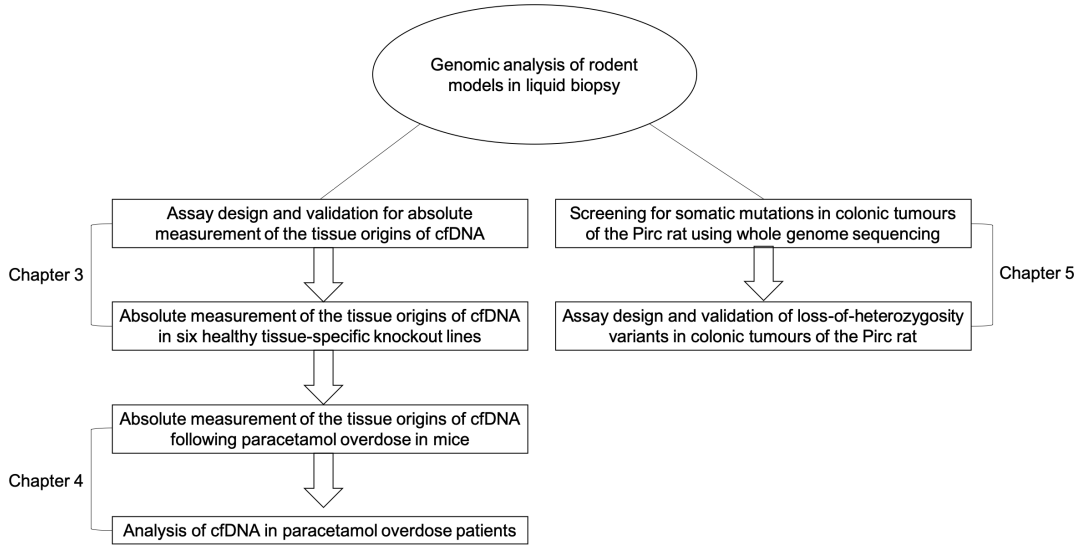

Figure 1.6.1 Flowchart of the study design.

Chapter 2 Materials and Methods

2.1 Rodent Samples

2.1.1 Mouse breeding

Mouse breeding was carried out in individually ventilated cages (4-6 mice per cage depending on litter size) with constant access to food and water in the Biomedical Research Facility (BRF), University of Edinburgh (UoE) under Home Office project license PPL P1070AFA9. Myeloid (LysM Cre) [73], lymphoid (hCD2-iCre) [74], cardiomyocyte (cTnT Cre) [75], hepatocyte (Alb Cre) [76] and striated muscle (MCK Cre) [77] Cre mice were obtained from the Jackson laboratory (Bar Harbor, Maine) in C57BL/6 background. Erythroid (EpoR Cre) [78] mice were obtained from Prof. Stuart Orkin (Cooperative Centers of Excellence in Hematology, Bostons Children's Hospital). These Cre mice were bred with mice carrying floxed *mT/mG* dual reporter gene (**Figure 2.1.1.1**) [79] was generated in C57BL/6 background to obtain cell/tissue-specific reporter F1 mice. Genotyping of F1 litter was performed by Transnetyx (Cordova, Tennessee, USA) on ear notch samples to confirm the presence of Cre recombinase and the floxed *mT/mG* gene. Male tissue-specific reporter mice (F1, n > 10 for each mouse line), that were heterozygous for Cre recombinase and the floxed *mT/mG* gene, were culled at 10-12 weeks old via exposure to gradually increasing concentrations of carbon dioxide gas (schedule 1 method) at 10% displacement rate (chamber volume / minute). Carbon dioxide flow was continued for 1-2 minutes after cessation of breathing.

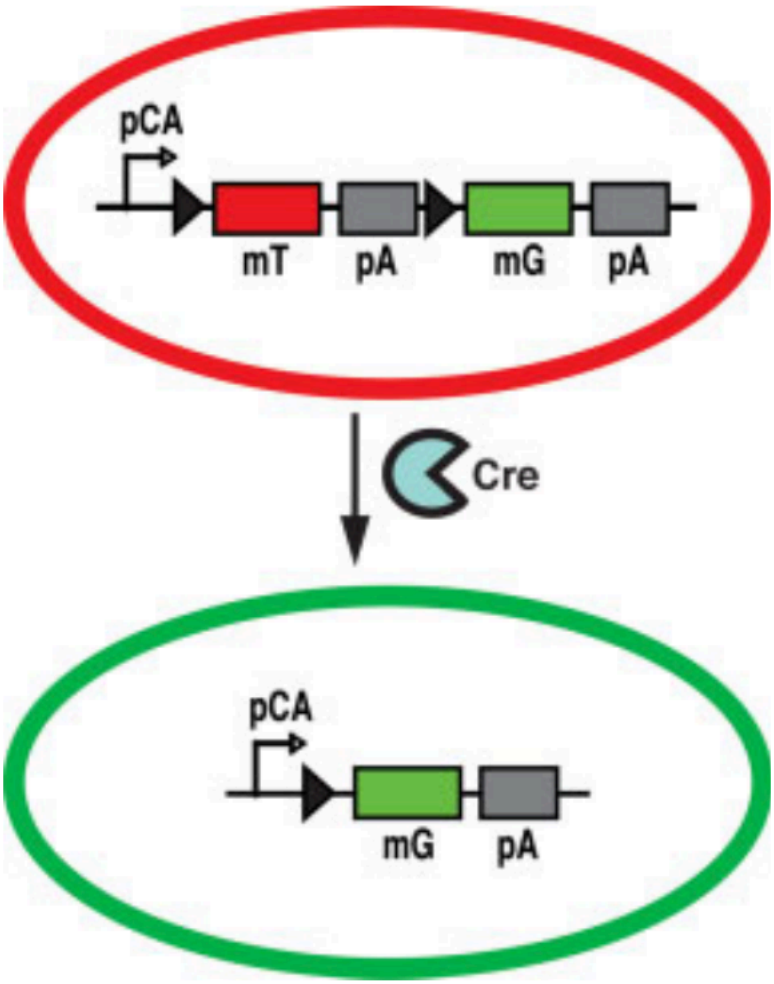

Figure 2.1.1.1 A schematic diagram of the mT/mG reporter gene before and after recombination.
The *mT/mG* reporter gene is driven by chicken beta-actin core promoter (pCA). Prior to Cre recombination (top), the *mT/mG* reporter gene contains a membrane-targeted mTomato (mT) flanked by two loxP sites (shown by black tiangles) and a membrane-targeted enhanced green fluorencent protein (mG). Both mT and mG are followed by a polyadenylation sequence (pA). Following Cre recombination (bottom), mT and its polyadenylation sequence are removed, leaving a loxP site and mG. The illustration is taken from [79].

2.1.2 Induction of APAP overdose in mice

Induction of paracetamol (APAP) overdose in mice were carried out under the Home Office project license PPL 70/7847 (Prof. Stuart Forbes, UoE). To induce APAP overdose in mice, mice were injected intraperitoneally (IP) with APAP at 350 mg/kg mouse body weight, then maintained at 30°C while given

semi-solid food for a certain period of time, and finally culled via exposure to increasing carbon dioxide (schedule 1 method) as previously described in Section 2.1.1.

To investigate the timepoint where cfDNA concentration is highest following APAP overdose, culling and plasma collection were performed at 8, 24 and 48 hours following APAP injection in C57BL/6 mice (n = 4 per time point). Negative control mice injected with saline were included in each experiment.

Following investigation in C57BL/6 mice, APAP overdose was induced in hepatocyte-specific gene reporter mice (n = 4). Culling and plasma collection for hepatocyte-specific reporter mice were performed at 8 hours following APAP dosing.

2.1.3 Rat Samples

Rat samples, including frozen and paraffin-embedded colonic tissues, were obtained from Dr. James Amos-Landgraff (University of Missouri, USA). Rats in this study were all female, inbred, aged between 77 to 276 days on August Copenhagen Irish (ACI) background strain (**Table 2.1.3.1**). The genotypes of the rats consisted of two rats are wildtype *Apc*$^{+/+}$ and 21 rats *Apc*$^{Pirc/+}$.

Table 2.1.3.1 Genotype and age of rats used in the study

Animal Number	Strain	Genotype of Apc (Pirc or WT)	Sex	Age (Days)
7	ACI	WT	F	247
8	ACI	Pirc/+	F	247
9	ACI	Pirc/+	F	247
10	ACI	Pirc/+	F	276
12	ACI	WT	F	276
13	ACI	Pirc/+	F	77
36	ACI	Pirc/+	F	77
37	ACI	Pirc/+	F	77
38	ACI	Pirc/+	F	77
39	ACI	Pirc/+	F	77
317	ACI	Pirc/+	F	188
390	ACI	Pirc/+	F	184
391	ACI	Pirc/+	F	184
393	ACI	Pirc/+	F	184
521	ACI	Pirc/+	F	156
526	ACI	Pirc/+	F	156
935	ACI	Pirc/+	F	220
943	ACI	Pirc/+	F	219
944	ACI	Pirc/+	F	219
945	ACI	Pirc/+	F	219
946	ACI	Pirc/+	F	219
1006	ACI	Pirc/+	F	161
1014	ACI	Pirc/+	F	161

2.2 Human Samples

Serum was collected from healthy volunteers (n = 11) and APAP overdose patients (n = 8) from the SNAP Clinical Trial [80] by Dr. James Dear (UoE). The overdose group was comprised of two groups: normal alanine aminotransferase (ALT, n = 4) and raised ALT (n = 4), indicating clinical/biochemically apparent liver injury. The protocol was approved by the Scotland A Research Ethics Committee, UK (ref no 10/MRE00/20).

2.3 Sample pre-processing

2.3.1 Collection and processing of tissue samples

Tissue samples were collected from mice post-mortem and stored at -70°C prior to analysis. Some tissue samples were incubated overnight in 40ml of formalin, and ethanol 70% until they were processed to be paraffin-embedded in the Pathology department, UoE. Paraffin-embedded tissues are stored at room temperature prior to analysis.

2.3.2 Collection and processing of blood samples

Blood samples were collected from mice post-mortem via the inferior vena cava using 1ml syringes with a 25G needle into EDTA tubes, and then mixed by inverting tubes for 3-5 times. Plasma was obtained by centrifugation of blood at 1000g for 10 minutes, then 16000g for 5 minutes, within 3 hours of blood collection. Mouse blood samples were processed at room temperature. Plasma samples were stored in -70°C prior to analysis.

Blood from clinical samples were centrifuged at 1000*g* for 15 minutes at 4°C and the supernatant was separated into aliquots and stored at -70°C prior to analysis. The protocol was approved by the Scotland A Research Ethics Committee, UK (ref no 10/MRE00/20).

2.3.3 gDNA extraction

gDNA was extracted from frozen tissue samples using DNEasy blood and tissue kit (QIAGEN) according to manufacturer's instruction. gDNA was stored in -20°C prior to further analysis.

2.3.4 Cell-free DNA extraction

cfDNA was extracted from plasma or serum samples using the QIAamp circulating nucleic acid kit (QIAGEN) according to the manufacturer's instructions. cfDNA samples were stored in -20°C prior to analysis. cfDNA samples were eluted from column using 50μl of elution buffer.

2.3.5 Bisulphite conversion of DNA and cfDNA

Bisulphite conversion of DNA and cfDNA was performed with the Epitect DNA bisulphite conversion kit (QIAGEN) using protocol for low input fragmented DNA. Bisulphite converted DNA was eluted from the QIAGEN column using 50μl of elution buffer.

2.3.6 Quantitation of DNA

gDNA samples extracted from tissues were quantified using Qubit dsDNA Broad Range assay (ThermoFisher Scientific) according to the manufacturer's instruction. Bisulphite converted DNA was quantified using Qubit ssDNA assay.

2.4 Histological analysis of tissue samples

2.4.1 Histological analysis of paraffin embedded tissue samples

Sections of 5μm were cut from paraffin-embedded tissues using a microtome. These sections were put on a slide and stained with hematoxylin and eosin (H&E). Sectioning and staining were performed by Dr. Helen Caldwell in the

Pathology Department, UoE. Slides were visualised using NDP Nanozoomer slidescanner.

2.4.2 Histological analysis of frozen tissue samples

Fluorescence microscopy of frozen tissue sections from reporter mice was performed to visually confirm Cre recombination in tissues, based on expression of cell membrane-localised enhanced green fluorescent protein (mG) and cell membrane-localised tdTomato (mT). Sections of 5μm were obtained from frozen tissues using a cryostat at 4°C. These sections were put on a slide with mounting medium prolong gold and DAPI to stain nuclear DNA and left overnight at 4°C. Sectioning and staining were performed by Dr. Helen Caldwell in the Pathology Department, UoE. Slides were visualised using a Britemac epifluorescence microscope in the Institute of Genetics and Molecular Medicine (IGMM) Advanced Imaging facility, UoE, with non-fluorescent frozen tissue slides from C57BL/6 mice as a negative control for tissue autofluorescence.

2.5 Polymerase chain reaction

2.5.1 Quantitative polymerase chain reaction (qPCR)

qPCR assays were performed on the beta-actin (*ACTB* / *Actb*) gene [45, 81] with SYBR Green to measure total concentration of cfDNA from mice and clinical samples. A standard curve of human or mouse gDNA was used in quantification. The standard curve was generated through serial dilution and ranged from 0.005 to 25ng/μl.

2.5.2 Droplet digital polymerase chain reaction

2.5.2.1 ddPCR for measurement of tissue origins of cfDNA

A probe-based ddPCR assay was designed to specifically quantify the number of 1lox and 2lox *mT/mG* alleles. The assay was adapted from a probe-based qPCR assay [82] and designed using Primer3 [83] following the Biorad ddPCR application guide, with a maximum amplicon size of 130bp to accommodate for small fragments of cfDNA. The assay was validated by amplification of gDNA containing only 1lox or 2lox DNA fragments by ddPCR (**Table 2.5.2.1.1**).

Table 2.5.1.1.1 Oligonucleotides for amplification of 1lox and 2lox mT/mG gene.

Oligonucleotide Name	Primer Sequence
mTmGF1	5'-GCCCTCGACACTAGTGAACC-3'
mTmGF2	5'-GGTAAGTAAGCTTGGGCTG-3'
mTmGR	5'-CTGCGGTCTTGGAGAAAC-3'
mTmGP	5'-[FAM]CGGATCATCACCGCGGATGGGTTGC[BHQ]-3'

1lox assay consisted of mTmGF1, mTmGR and mTmGP. 2lox assay consisted of mTmGF2, mTmGR, mTmGP. FAM: 6-carboxyfluorescein; BHQ: black hole quencher.

An input of 10ng of cfDNA was required for each well of ddPCR. Amplification of 1lox and 2lox alleles were performed in separate wells. Each sample was measured at least in triplicate. The amplification was run on a thermal cycler as follows: 10 min of activation at 95°C, 40 cycles of a two-step amplification protocol (30s at 94°C denaturation and 60s at 60°C), and a 10-min inactivation step at 98°C.

2.5.2.2 ddPCR for measurement of liver contribution in human cfDNA

ddPCR was performed to measure the contribution of liver tissue to cfDNA in clinical samples using a methylation-specific ddPCR assay [84] on 10ng of bisulphite-converted cfDNA. The assay consisted of a mixture of primer pairs and probes (**Table 2.5.2.2.1**) that specifically amplified a methylated (from liver)

and unmethylated (from other tissues) region in the *PTK2B* gene (chr8: 27,183,116-27,183,176, hg19). The amplification reaction was run on a thermocycler as follows: 10 min of activation at 95°C, 45 cycles of a two-step amplification (94°C for 15s and 60°C for 60s), and a 10-min inactivation step at 98°C.

Table 2.5.2.2.1 Oligonucleotides used in liver-specific methylation assay.

Assay	Primer/Probe	Sequence
Methylated	Forward	5'-TTTATTTGTTCGGTCGATTTATTTGTA-3'
	Reverse	5'-AACGACTACTCCTCGAACACCG-3'
	Probe	5'-[FAM]-TTGTCGTCGTTTCGGTT-[BHQ]-3'
Unmethylated	Forward	5'-TGTATATTTATTTGTTTGGTTGATTTATTTGTA-3'
	Reverse	5'-CCAACAACTACTCCTCAAACACCA-3'
	Probe	5'-[VIC]-TTTGTTGTTGTTTTGGTTTA-[BHQ]-3'

2.5.2.3 ddPCR for detection of LOH in rat samples

Probe-based ddPCR assays were designed across rat chromosome 18 to validate the presence or absence of LOH. Dual-labeled probes (FAM and HEX) were used for each assay. FAM probes specifically bind to different allele of a heterozygous variant compared to HEX probes. ddPCR assays were designed using Primer3 [83] following the Biorad ddPCR application guide, with maximum amplicon size 130bp. The amplification was run on a thermal cycler as follows: 10 min of activation at 95°C, 40 cycles of a two-step amplification protocol (30s at 94°C denaturation and 60s at a specific annealing/extension temperature on **Table 2.5.2.3.1**), and a 10-min inactivation step at 98°C.

Table 2.5.2.3.1 Oligonucleotide sequences of ddPCR assays for screening of LOH.

Genomic Coordinate	Primer F	Primer R	Probe [FAM/HEX]	Annealing temperature (°C)
chr18:27100213	GTCAGGAAGACGACTATG	GCTGTTCTTCCTCAGA	TGAC[A/T]AGCCTACCAACTACAGTGAACGTT	58.1
chr18:27043184	CAGTCCAGACTGTTGCTAGA	GGCTCTGGGTTTTCTTTACAA	CAGCAACA[C/T]ATAAAAGAATACCCAAACA	52.7
chr18:27112121	GAGAGACACATCGCATACAT	GGGTTTTGCTGACAAAT	ACATTATGTCTATG[C/T]ATATATGACAAATT	52.7
chr18:28062387	ACTGTGAGCCCTTTTTCTCC	CCACCTCAGCACAGGTAATC	TG[G/T]AATGCCTTTGTGTTTCCACTT	62
chr18:25461411	CCTGAGATCTACTTGCTC	GCCATCAGTGGATGACTA	TGGACACAA[A/T]AGAGCTCTCTGAAGGAACG	62
chr18:87928418	TGCTTAATGCAGTGGTCTCC	GTCCTGGTACACATACAC	AG[T/A]CTTAATTCAATCACTTCTGTGTCTGGA	61.2

2.6 Fragment analysis of cfDNA

Fragment analysis of cfDNA was performed using Agilent DNA Bioanalyser high sensitivity chip at the Welcome Trust Clinical Research Facility.

2.7 Analysis of protein biomarkers from plasma and serum samples

Liver function tests (alanine aminotransferase / ALT, aspartate aminotransferase / AST, albumin (ALB), bilirubin, and glutamate dehydrogenase / GLDH) were measured from mouse plasma and human serum samples by Dr. Forbes Howie at the Shared University Research Facility, UoE.

2.8 Absolute measurement of the tissue origins of cfDNA

Absolute measurement of the tissue origins of cfDNA in mice was achieved by analysing cfDNA from F1 tissue-specific reporter mice. cfDNA originating from the cells of interest, which express Cre recombinase from a cell-specific promoter, were recombined (1lox), as opposed to unrecombined (2lox) from other cell types in the body that do not express Cre recombinase (**Figure 2.8.1**). Analysis of Cre recombination in gDNA and cfDNA of tissue-specific reporter mice was performed to show successful recombination in mouse tissue and the contribution of the corresponding Cre-driven cell/tissue types to cfDNA. The metric used to show Cre recombination was percentage recombination (1lox%).

$$1\text{lox}\% = \frac{total\ 1lox\ allele}{1lox+2lox\ allele} * 100\%$$

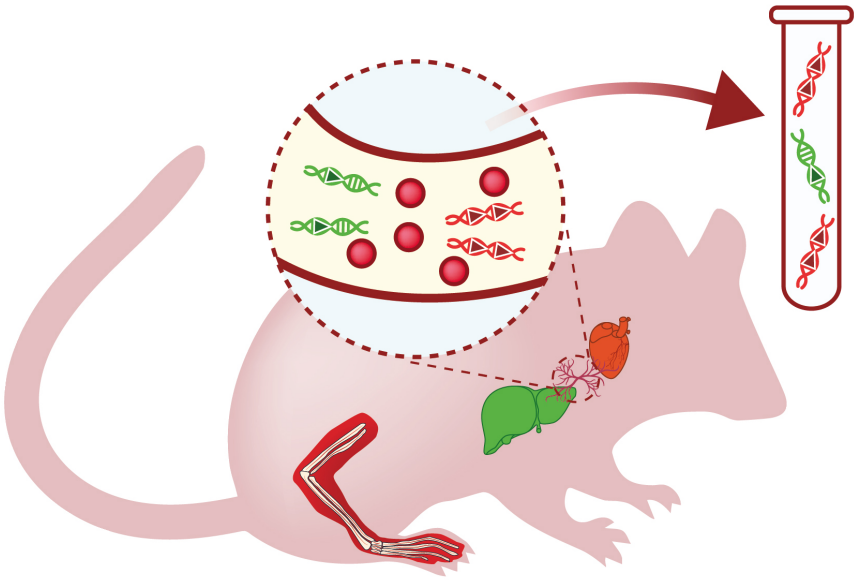

Figure 2.8.1 A schematic representation of the overall strategy for absolute quantitation of the tissue origins of cfDNA.
A conditional deletion is generated in the tissue of interest (e.g. liver) for each mouse model, containing a tissue/cell-specific Cre recombinase and the reporter gene floxed mT/mG. Cre recombination occurs in target cells/tissues, causing deletion of a loxP site and the mT gene, leaving one loxP site in the DNA (1lox), and enabling the expression of the mG gene (green-colored) as opposed to other tissues in the body that contain two loxP sites (2lox, colored red). Histology and DNA analysis are performed in mouse tissues to confirm Cre recombination. cfDNA, containing a mixture of 1lox and 2lox alleles released by various tissues through normal tissue turnover, can be extracted and analysed to reveal absolute contribution of cell-type/tissue of interest.

2.9 Whole genome sequence analysis

2.9.1 Whole genome sequence generation and mapping to reference genome

gDNA extracted from normal liver tissue of a ACI wildtype rat (*Apc*$^{+/+}$), and paired liver-colonic tumour tissue of two ACI-Pirc rats (*Apc*$^{Pirc/+}$). DNA samples were sent to Edinburgh Genomics for library preparation with the Illumina TruSeq Nano kit followed by sequencing on the Illumina HiSeqX platform at 30x and 60x average coverage for liver and colonic tumour DNA respectively. The raw sequence data was delivered in fastq file format and analysed

alongside an F344N/Crl rat genome publicly available [85]. The fastq files were mapped to Brown Norway rat (*Rattus norvegicus*) reference genome from the UCSC genome browser (RGSC Rnor_6.0) [86, 87]. Pre-processing of sequence data to obtain analysis-ready alignment files (.bam) was performed according to the best practices from the Genome Analysis Toolkit (GATK) [56].

2.9.2 Detection and validation of somatic LOH

Analysis ready alignment files were processed with HaplotypeCaller to obtain a file containing germline variants (.vcf) according to the best practices from GATK [56]. To screen for somatic LOH, variant files were filtered based on genotyping quality (GQ) above 30, and per sample depth above 20. The analysis was also restricted to biallelic variants that passed GATK default hard-filtering criteria. Chromosome, genomic coordinate, and allele depth for each sample was extracted from the filtered variant file using bcftools. Alternate allele balance for every sample was calculated from the allele depth data. Alternate allele balance values were plotted for each paired liver-colonic tumour sample per chromosome to screen for deviation in alternate allele balance of the tumour sample from the germline liver sample.

To approximate the beginning and ending of LOH, circular binary segmentation analysis [88] was performed using an R package, called PSCBS [88], on a metric describing the allelic shift between the normal and tumour sample, called the pseudo odds:

$$\text{pseudo odds} = \frac{(nrd+1)/(nad+1)}{(trd+1)/(tad+1)},$$

nrd referred to the depth of reference allele in a normal sample, nad referred to the depth of alternate allele in a normal sample, trd referred to the depth of reference allele in a tumour sample, and tad referred to the depth of alternate allele in tumour sample.

2.9.3 Detection of somatic single nucleotide variants (sSNVs) and small insertion and deletion (indel) variants

Somatic single nucleotide variants (sSNVs) and small indel variants were detected by using three different somatic variant callers, namely Mutect2 [57], Varscan2 [89], and Strelka2 [90]. A confident list of sSNVs and small indels for each sample were obtained by overlapping sSNVs and small indels from these three tools. The confident list contained variants detected by at least two out of three somatic variant callers. Variants in the confident list were annotated by using snpEff [91] and then compared to the cancer gene census project from the COSMIC database for human cancers [92, 93].

2.9.4 Detection of structural and copy number variants

Structural and copy number variants, including deletions, duplications, and inversions, were detected by using two somatic structural variant callers, namely Lumpy [94] and Manta [95]. A confident list of structural variants was obtained by overlapping variants obtained from both tools and further filtered based on allele balance (greater than 0.15 for tumour sample, less than 0.1 for normal sample) and number of paired-end reads supporting evidence (more than 10 supporting evidence). Variants were annotated by using

ANNOVAR [96]. A list of genes from exonic variants were compared to the cancer gene census project from the COSMIC database for human cancers [92, 93].

2.10 Statistics

Normality of data distribution was analysed using Shapiro-Wilk test. Comparison between two independent group was performed using Mann-Whitney U test. Comparison between 3 or more groups were performed using Kruskall-Wallis test. Associations between datasets were analysed using Spearman rank correlation test. A more detailed statistical analysis is available in Appendix C.

Chapter 3 Absolute measurement of the tissue origins of cell-free DNA using healthy tissue-specific reporter mice

3.1 Introduction

Studies of the tissue origins of cfDNA showed that cfDNA predominantly originates from hematopoietic tissues [14, 23–25], with minor contributions from liver and endothelial cells [26]. These studies were performed in human mainly by comparing DNA methylation patterns or nucleosomal profiling of different tissues in cfDNA. Analysis of the tissue origins of cfDNA using methylation pattern relies on distinct DNA methylation signatures found in different cell types and tissues. These distinct methylation patterns can be observed in cfDNA by targeted or whole-genome bisulphite sequencing. Targeted bisulphite sequencing of cfDNA using a specific methylation marker may reveal the contribution of a specific tissue [24, 84], whereas whole-genome bisulphite sequencing may reveal the contribution of many tissues of known methylation signature [25, 26]. The contribution of liver and white blood cells to cfDNA was first discovered using whole-genome bisulphite sequencing approach, whereas erythroid cells using targeted bisulphite sequencing approach [24, 25].

Analysis of the tissue origins of cfDNA using nucleosomal profiling exploits the variation of nucleosome positionings between cell types. The boundaries of cfDNA fragments correlate with cell type-specific nucleosome positioning, and the fragmentation of cfDNA contains evidence of the epigenetic landscape of the cells of origin. The tissue origins of cfDNA can be estimated by deep

sequencing of cfDNA and comparison of its nucleosomal profile with existing nucleosome profiles of human cells. Similar to findings from analysis of DNA methylation pattern in cfDNA, analysis of nucleosomal profiling in cfDNA showed that blood cells, specifically myeloid and lymphoid, are the major contributors to cfDNA [14].

There is a substantial discordance between estimates of tissue contribution generated by different analytical methods in human. Contributions of tissues generated by these studies were presented as relative measures as opposed to absolute [1]. Further, comparisons of tissue origins of cfDNA with DNA methylation patterns or nucleosomal profiling require several steps to be performed, such as DNA bisulphite conversion and sequencing library preparation, prior to the actual measurement of tissue origins. This multi-step approach may contribute to the variation of estimates between these studies.

Study of tissue origins of cfDNA in animal models has never been performed. Measurements of tissue origins of cfDNA in animal models can provide an absolute measurement due to the possibility of generating specific genetic alterations in tissue of interest using Cre-loxP recombination system [97]. Further, control of different factors, that may contribute to experimental variability, such as diet and age, is easier in animal models than in human. The subsequent sections of Chapter 3 describe the utilisation of tissue-specific reporter mouse models to measure the tissue origins of cfDNA, from generation of mouse lines, establishment of measurement assays, and estimation of tissue origins of cfDNA in healthy mice.

3.2 Generation of conditional deletion in tissue-of-interest

In this project, a specific deletion in tissues of interest was generated using Cre-loxP recombination system in mice [97]. Mice homozygous for the floxed *mT/mG* gene (see section 2.1.1) were crossed with mice carrying Cre recombinases in tissues of interest. Recombination were expected to occur only in the mT/mG gene, and not in other loci in the genome. Mice born from this cross (F1) were genotyped to confirm heterozygous genotype for both the *mT/mG* gene and a specific Cre-recombinase. Across six tissue-specific reporter mouse lines generated, a minimum of 10 mice were successfully obtained for each line. A comparison of mouse body weight at 10-12 week old between reporter mouse lines (**Figure 3.2.1**) showed that hepatocyte-specific reporter mice (*AlbCre*$^{+/-}$; *mT/mG*$^{+/-}$) are significantly heavier than control mice without Cre recombinase (*mT/mG*$^{+/-}$). The median of body weights from other mouse lines are not significantly different from control. Despite showing statistically significant difference in body weight, median difference between hepatocyte-specific reporter mice and no Cre controls was only 3.3gram. Tissue-specific reporter mice showed no adverse phenotype, consistent with the previously reported phenotype and with the absence of change in coding sequence in any of the parental line genes [73–78].

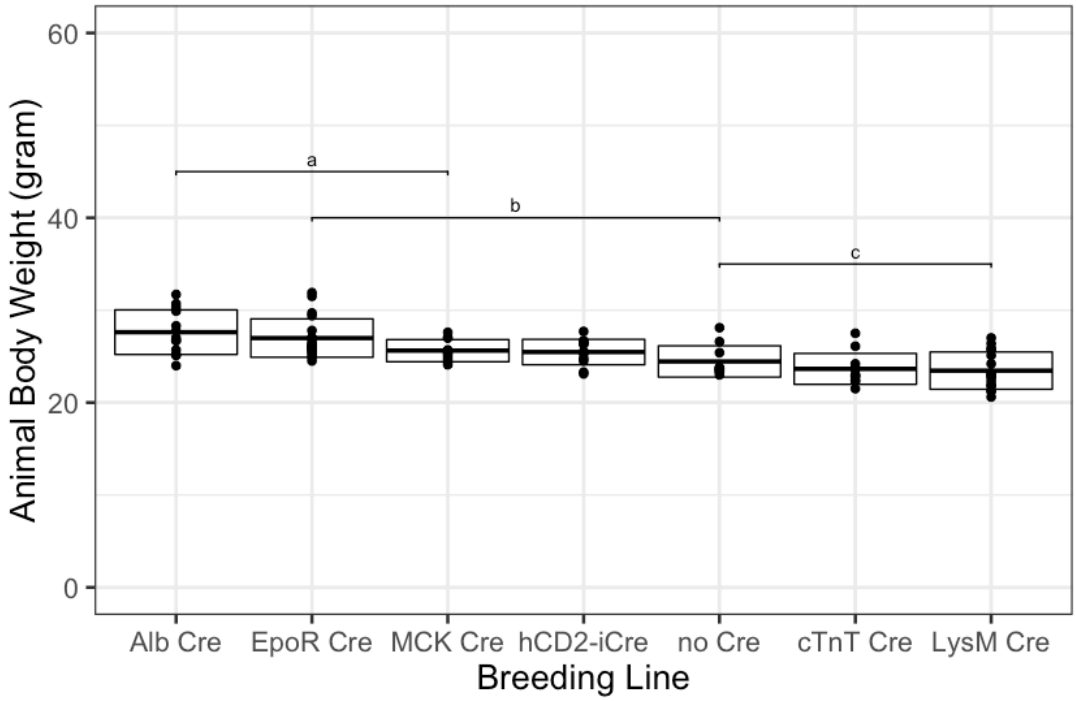

Figure 3.2.1 Comparison of body weight from different tissue-specific reporter mouse lines at 10-12 week old.
Statistical analysis was carried out using Mann-Whitney U tests separated mouse lines into three groups shown by letters above the boxplot (a, b or c). The median body weight of mice from the Alb Cre line was significantly different to no Cre controls.

3.3 Design and validation of assay to detect conditional deletion in genomic and cell-free DNA

A ddPCR assay was designed and validated to quantify recombination in the floxed *mT/mG* gene. The assay consisted of a combination of primer pairs and probe that specifically amplify the 1lox or 2lox alleles of the *mT/mG* gene. Amplification of gDNA was performed to demonstrate specificity for each target allele (**Figure 3.3.1a and b**). Amplification of 2 lox gDNA with the 1lox assay showed no positive droplet (**Figure 3.3.1a**). Amplification of the 1lox gDNA with the 2lox assay showed 13 positive droplets out of a total of 16467 (less than 0.1%) (**Figure 3.3.1b**). Assay sensitivity was tested using a dilution series of target alleles to show that the assay reproducibly quantifies less than 30

target alleles (**Figure 3.3.1c**). Non-specific amplification was not observed in C57BL/6 background gDNA, nor in the water negative control.

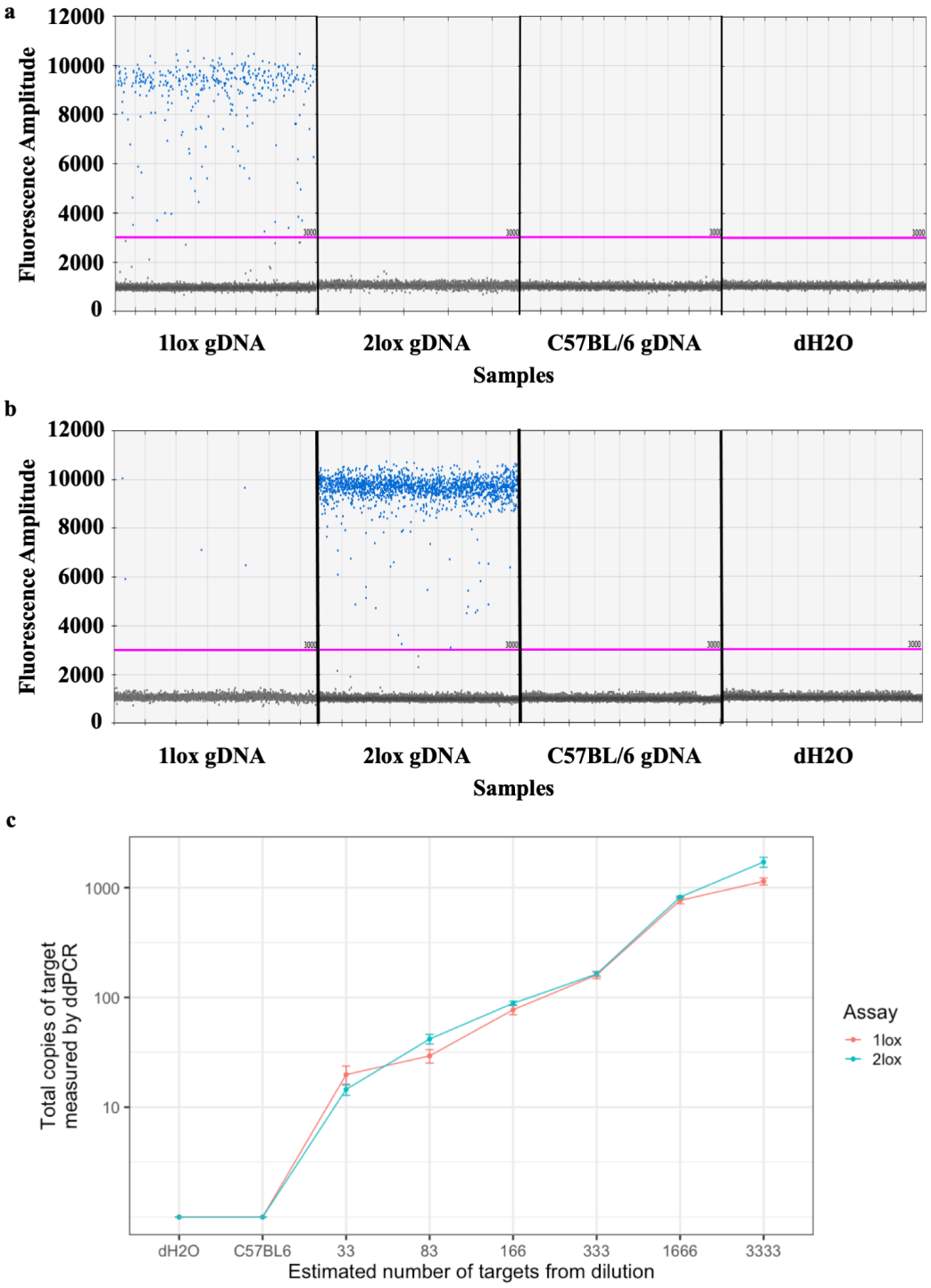

Figure 3.3.1 Validation of ddPCR assays on gDNA containing 1lox or 2lox alleles.
(**a**) 1lox ddPCR assay specifically amplified gDNA containing 1lox alleles, (**b**) 2lox ddPCR assay amplified gDNA containing 2lox alleles with less than 0.1% of positive droplets from 1lox gDNA. The pink line shows the threshold between positive and negative droplets at fluorescence amplitude 3000. (**c**) Dilution experiment showed both assays robustly detected less than 30 targets. Each dot showed mean of 1lox or 2lox copies measured by ddPCR assays in six replicates. Error bars showed standard error.

3.4 Determination of DNA input requirement and minimum number of mice to analyse conditional deletion

Aside from assay sensitivity and limit of detection explained in section 3.3, the requirement of DNA input into ddPCR assay was determined based on the amount of cfDNA that can be extracted from individual mice, and a theoretical Poisson probability distribution of the corresponding input amount in ddPCR platform (**Figure 3.4.1**). An input of 10ng into ddPCR allowed for triplicate measurement of cfDNA with 1lox/2lox ddPCR assay and high theoretical measurement sensitivity.

An experiment in C57BL/6 mice showed that less than 5ng of cfDNA can be extracted from a mouse, and less than 10ng from a pool of 5 mice. More than 70ng of cfDNA can be obtained from a pool of 10 mice (**Figure 3.4.1a**). This indicated less efficient recovery of cfDNA during extraction protocol using a lower volume of plasma.

The incorporation of DNA template into 20000 droplets during ddPCR occurs randomly. However, statistics shows that this random incorporation follows the Poisson distribution statistics. A theoretical Poisson probability distribution of 10ng input DNA (approximately 3333 targets) showed that majority of positive droplets had one target molecule per droplet (**Figure 3.4.1b**). This strategy in ddPCR, whereby a certain amount of DNA input is used to ensure that a majority of positive droplets contain only 1 copy of target, is called the limiting dilution strategy.

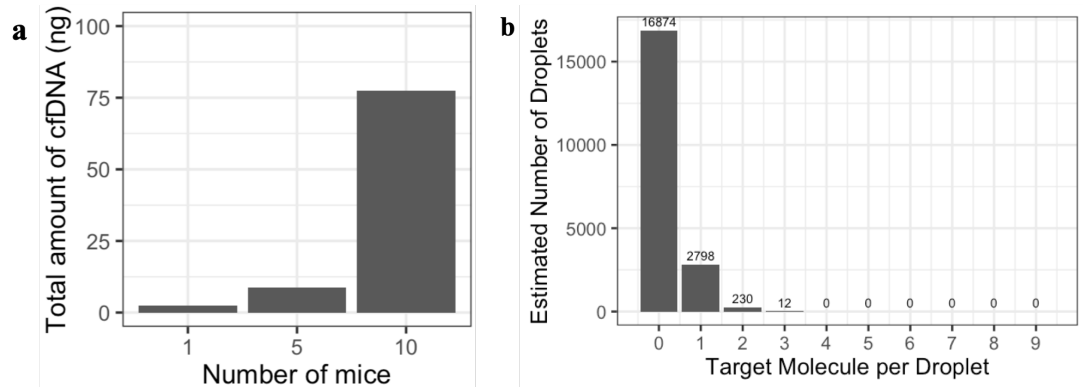

Figure 3.4.1 Determination of ddPCR DNA input requirement.
DNA input requirement depends on amount of cfDNA that can be extracted from mice and optimum Poisson probability distribution in ddPCR platform to achieve limiting dilution. (**a**) Total amount of cfDNA that can be extracted from one, five and ten pooled C57BL/6 mice. (**b**) A Poisson probability distribution of 10ng input DNA into ddPCR platform (approximately 3333 input target molecule).

3.5 Analysis of conditional deletion in gDNA from mouse tissues

Cre recombination was confirmed in expected tissues using ddPCR (**Figure 3.5.1**) and fluorescence microscopy (**Figure 3.5.2**) based on the expression of the *mT/mG* reporter gene prior to analysis of cfDNA. To check for specificity of Cre recombination, amplification of gDNA extracted from 16 tissue types for each mouse line with 1lox/2lox ddPCR assay was performed. Percentage recombination of different tissues of reporter mice was calculated from the number of 1lox and 2lox alleles in corresponding tissues. Highest specificity was found in hepatocyte and cardiomyocyte reporter lines, where recombination was observed in liver and heart, respectively, with minimal recombination in other tissues (**Figure 3.5.3**). Specificity of Cre recombination was more variable, but generally lower in the other reporter lines. Cre recombination can be observed across tissues of myeloid-specific reporter mice. This was likely due to the presence of resident macrophages and

difficulty of exsanguinating blood completely from analysed tissues. Lymphoid-specific reporter mice showed the least sensitivity among the six lines analysed in the study.

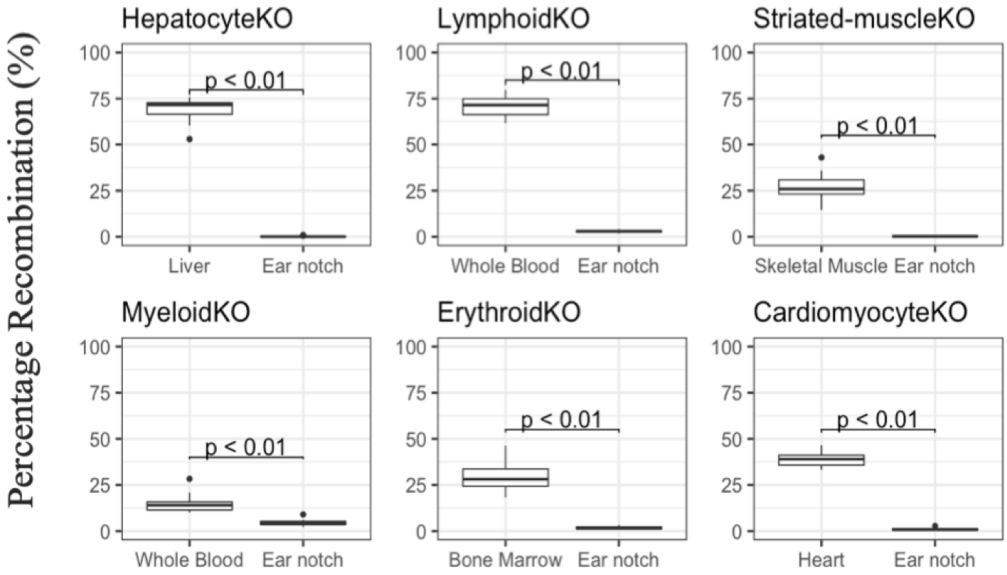

Figure 3.5.1 Confirmation of Cre recombination in tissue-specific reporter mouse lines using ddPCR.
1lox and 2lox alleles were measured using ddPCR in tissues where recombination was expected vs. not expected for each reporter lines (n > 10 mice for each line). p-value obtained from a Mann-Whitney U test.

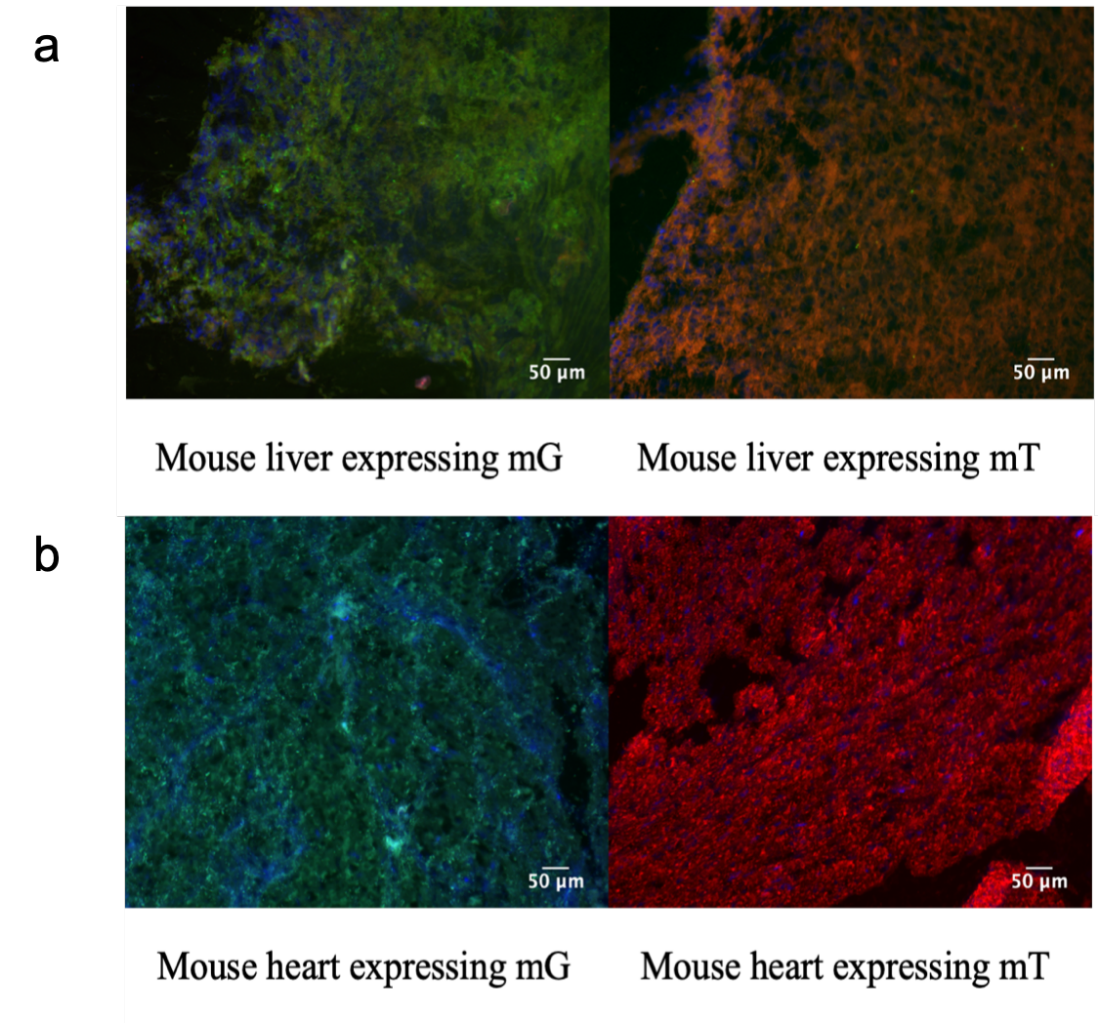

Figure 3.5.2 Visualisation of Cre recombination in liver and heart tissue sections of hepatocyte and cardiomyocyte-specific reporter mice.
(a) liver of a *AlbCre*;*mT/mG* mouse, liver of a *mT/mG* mouse. (**b**) heart of a *cTnTCre*;*mT/mG* mouse, heart of a *mT/mG* mouse. Green colour in the tissues derived from the expression of mG, and red from mT

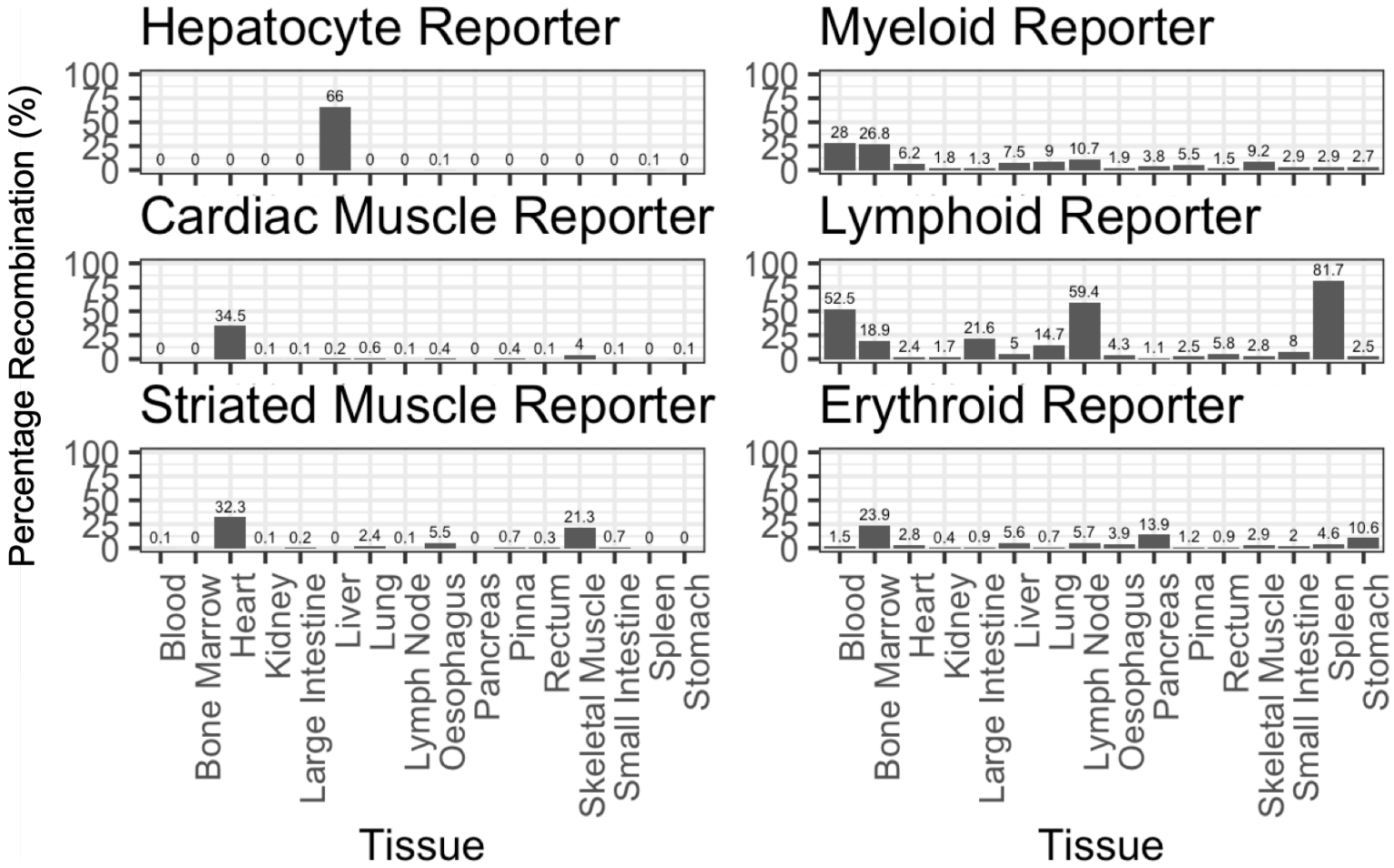

Figure 3.5.3 Specificity of Cre recombination in tissue-specific reporter mouse lines across tissues.
Measurements of 1lox and 2lox alleles using ddPCR across tissues showed high Cre recombination specificity in hepatocyte and cardiac muscle-specific reporter mice, and lower specificity in other reporter lines.

3.6 Analysis of conditional deletion in cell-free DNA of healthy mice

To determine the tissue origins of cfDNA, we analysed cfDNA from the six tissue-specific reporter lines using ddPCR. Plasma from 10 or more mice for each reporter line was pooled prior to cfDNA extraction, to account for the small circulating blood volume of individual mice and the low physiological concentration of cfDNA. Sufficient DNA was extracted for triplicate measurements of the pooled sample from each line by ddPCR (**Figure 3.6.1a**). DNA fragment analysis confirmed the expected cfDNA fragment size profile which followed the mono-, di- and tri-nucleosome pattern associated with apoptosis, with larger cfDNA fragments most likely due to tissue necrosis and lysis (**Figure 3.6.1b**).

Contribution of different cell types to cfDNA was quantified by counting the number of 1lox and 2lox alleles in the cfDNA of mice for each reporter line and subsequently the percentage recombination (1lox%) for each of the lines was calculated (**Figure 3.6.1c**). Absolute measurement of cfDNA showed that myeloid, lymphoid and erythroid lineages were the major contributors to the levels of cfDNA in healthy mice, contributing 39.7%, 27.8% and 10.7% respectively, with a small contribution from hepatocytes (4.2%). Both cardiomyocyte and striated muscle cells had undetectable contribution to the levels of cfDNA (**Figure 3.6.1d**).

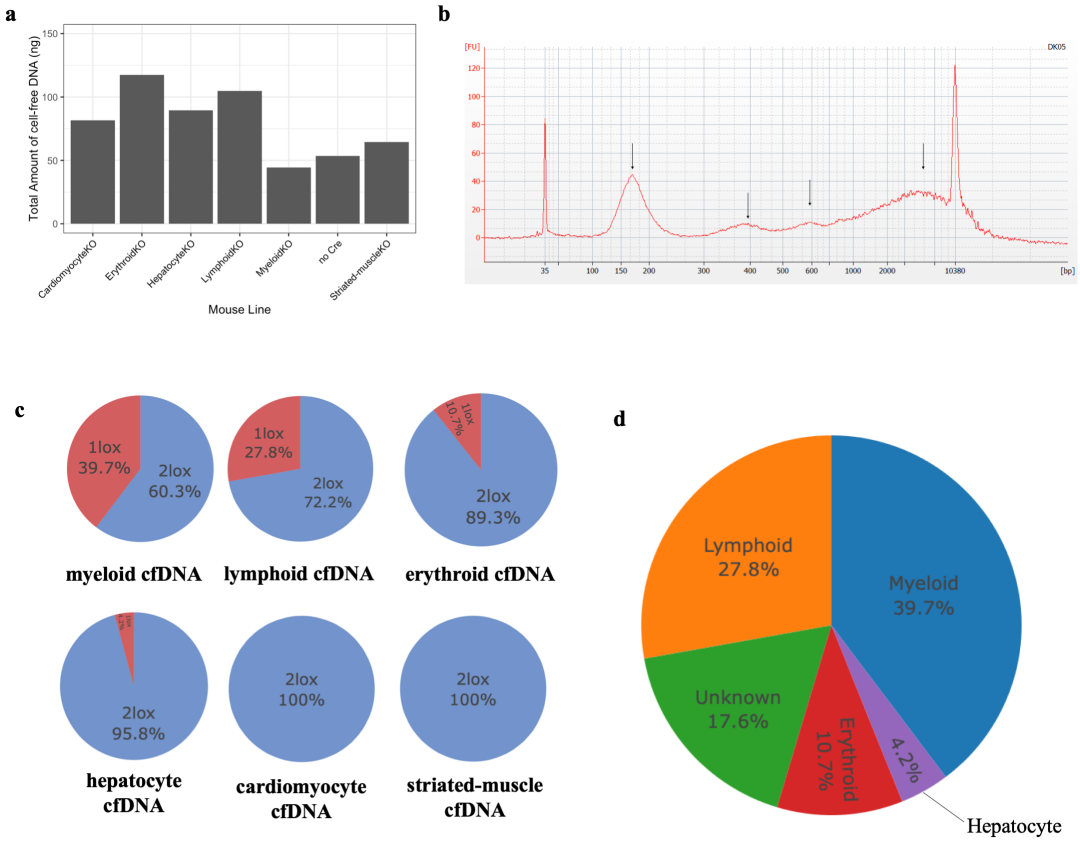

Figure 3.6.1 The tissue origins of cfDNA in healthy mouse models.
(**a)** Total amount of cfDNA obtained from a minimum of 10 mice (pooled) per reporter line measured by qPCR of *ACTB* gene. (**b**) Representative cfDNA fragment analysis using the Agilent DNA bioanalyser high sensitivity assay from hepatocyte-specific reporter mice. Fragments of cfDNA followed mono-, di-, and tri- nucleosome fragment sizes, and large cfDNA fragment size, shown by black arrows (**c**) Percentage recombination (1lox%) for each reporter mouse line showing contribution of different cell types to total cfDNA. (**d**) A cumulative percentage of cell/tissue contributions to cfDNA from ddPCR.

3.7 Discussion

This chapter demonstrates the utility of tissue-specific reporter mouse models as a platform to examine the tissue origins of cfDNA. Here, the first absolute, direct measurement of the tissue origins of cfDNA in healthy mice is presented. Experiments in mice show that hematopoietic cells, including myeloid, lymphoid and erythroid cells, are the major components of cfDNA in the healthy state, with a minor contribution from liver and no detectable contribution from cardiomyocytes and skeletal muscle cells. These results

show concordance with other indirect methods, such as nucleosomal profiling and methylation analysis, that have assessed the tissue origins of cfDNA [14, 24–26], although these studies show considerable variation in the estimates of tissue origin, which may in part be due to the indirect or relative measurement methods [1].

Aside from tissue contributions from hematopoietic, hepatocyte and muscle cells, an unknown tissue contribution of 17.6% is unaccounted. A recent investigation of the tissue origins of cfDNA in human using tissue-specific DNA methylation patterns revealed that vascular endothelial cells contribute 10% to cfDNA pool of healthy individuals [26]. The unknown contribution found in this study might be a mixture of contributions from vascular endothelial cells, technical variability, and possibly minor contributions from other cell types yet to be investigated.

Measurement in this study benefits from the ability to control external variables, further reducing sample variability. This is shown by low variability of body weight observed between mouse lines. This experimental design may therefore be more accurate and less error-prone than indirect relative methods.

In this project, the *mT/mG* reporter gene was chosen as the target allele deleted in tissues of interest. The *mT/mG* reporter gene was selected because its expression and deletion does not cause any harmful phenotype in mice. The *mT/mG* reporter gene is located in *ROSA26* locus in the mouse chromosome 6. It is an ubiquitously expressed locus, thus, eliminating variability from potential silencing of local chromatin structure between different cell types [79] and nucleosomal spacing in cfDNA. Additionally, the

utilisation of reporter gene *mT/mG* is beneficial because it allows for visual evaluation of Cre recombination in mouse tissues. Finally, ddPCR assay can be designed and validated around loxP sites of the 1lox and 2lox *mT/mG* reporter gene with high specificity and sensitivity.

In this project, the tissue-specific reporter mice used were heterozygous for the floxed *mT/mG* gene and Cre-recombinase. Despite having enough target alleles to be analysed in tissues and cfDNA, investigation of the tissue origins of cfDNA using tissue-specific reporter mice with a homozygous floxed *mT/mG* gene would be more advantageous than heterozygous. Mice with a homozygous floxed *mT/mG* gene would have double the number of target alleles to be analysed. This in turn would reduce the number of mice utilised for each mice (reduction principle of mouse study). In this project, a decision was made to utilise tissue-specific reporter mice with heterozygous floxed *mT/mG* gene as opposed to homozygous because of time constraints in breeding these homozygous mice.

In this study, ddPCR was selected as a platform to analyse cfDNA following isolation. Measurement of tissue origins of cfDNA using ddPCR can be performed directly following cfDNA extraction from plasma, without the need to process samples prior to analysis, such as DNA bisulphite conversion for comparison of methylation patterns, or library preparation for nucleosomal profiling in human. Additionally, with the right input DNA, measurement using ddPCR may provide high accuracy and sensitivity. Altogether, the utilisation of ddPCR to analyse the tissue origins of cfDNA allows for direct, accurate and sensitive measurement.

Despite having the advantage of high accuracy and sensitivity, analysis using ddPCR provides only quantitative information in assessing the specificity of Cre recombination in mouse tissues. Analysis of Cre recombination in tissues may also be complemented by visualisation of reporter genes in tissues using epifluorescence microscopy. This may provide spatial information and show whether Cre recombination in tissues occur in the respective tissues or in blood cells circulating in the tissues.

Aside from using ddPCR, measurement of floxed *mT/mG* alleles in tissues and in cfDNA may also be performed using sequencing. Measurement of floxed *mT/mG* alleles using sequencing may involve enrichment of target alleles using PCR primers or capture using hybridisation probe and library preparation for sequencing. 1lox and 2lox *mT/mG* alleles may be quantified from sequencing depths. This method requires more steps to be performed prior to quantitation of target alleles and will likely introduces more errors when compared to measurement using ddPCR.

Compared to studies in human, this study is limited in biological replicates. This limitation is caused by limited amount of blood that can be obtained from a single mouse. A plasma pool of at least 10 mice was necessary to obtain sufficient material for ddPCR assay measurement instead of measurement in 10 individual mice. This number of mice is carefully selected by taking into account the amount of cfDNA that can be obtained, as well as the sensitivity and specificity of ddPCR assay. A reduction in input cfDNA amount would compromise the ability of ddPCR assays in detecting tissues with low contribution to cfDNA. In addition, haemolysis and contamination of large

cfDNA fragments that arises from lysis of blood cells during sample processing may have led to an overestimation of the fraction of cfDNA derived from white blood cells, although blood cell lysis were minimal in all samples used in this study. Finally, the contribution of cells measured in this study is dependent upon the activity and specificity of Cre-recombinase, which varies between reporter lines. Due to non-specific Cre recombination in myeloid and lymphoid-specific reporter lines, the contribution of myeloid and lymphoid cells to cfDNA may be overestimated. Further, for each reporter lines, there may be contribution from other tissues which recombination status was not analysed, such as brain, skin, and gonad. This specificity issue is similar to investigation with other methodology such as analysis of tissue-specific methylation pattern, which utilises specific methylation markers of the tissue-of-interest against all other tissues in the body [24–26, 84].

Further experiments may be performed to show the contribution of cells/tissues that are yet to be analysed in this study, such as endothelial cells, gonads, as well as kidney, or to show reproducibility of tissue contributions found in this study. A similar tissue-specific reporter experiment may be performed using a different reporter gene as target allele, such as the Fucci2a reporter gene [98], or using a different Cre recombinase under a more specific promoter. Alternatively, measurement of the tissue origins of cfDNA may be possible using tissue-specific knock-in [99] to insert a sequence of DNA to be detected in cfDNA, instead of deletion of a reporter gene. Most importantly, a follow up experiment to validate findings from healthy reporter mice should be performed. An experiment whereby apoptosis or necrosis is induced in tissues

with low contribution to cfDNA would validate the robustness of tissue-specific reporter models in measuring the tissue origins of cfDNA. Apoptosis or necrosis of could also be induced by artificially injuring tissues of interest, a subject discussed further in Chapter 4.

Chapter 4 The effect of tissue injury on the tissue origins of cell-free DNA

4.1 Introduction

This chapter discusses the effect of tissue injury to the tissue origins of cfDNA in tissue-specific reporter mice. This investigation serves as a validation of the measurement of tissue origins of cfDNA in tissue-specific reporter mice.

4.1.1 Current understanding of the tissue origins of cell-free DNA following tissue injury

cfDNA is predominantly released into circulation through cellular death mechanisms, namely apoptosis and necrosis. It is, therefore, logical to assume that an increase of cell death from a specific tissue would increase the contribution of the respective tissue in circulation. Various experiments in human have confirmed that cfDNA fragments originating from damaged tissues can be detected in circulation. These experiments include the detection of cfDNA fragments from damaged tissues in patients receiving solid organ transplantation [100, 101], patients with myocardial infarction [33], patients with traumatic brain injury, and patients with pancreatitis [32]. These studies also confirmed the potential of cfDNA analysis as a biomarker of tissue injury in various pathological conditions.

4.1.2 APAP overdose and tissue injury

Paracetamol (N-acetyl-para-aminophenol, or APAP) is an analgesic drug that is safe when taken at therapeutic doses. APAP overdose is the most common cause of acute liver failure in USA and Europe. Toxicity of APAP (**Figure**

4.1.2.1) is caused by the generation of a metabolite, called N-acetyl-p-benzoquinone imine (NAPQI), in excess [102]. NAPQI is a reactive metabolite that reacts with cellular macromolecules and initiate cellular death [103, 104].

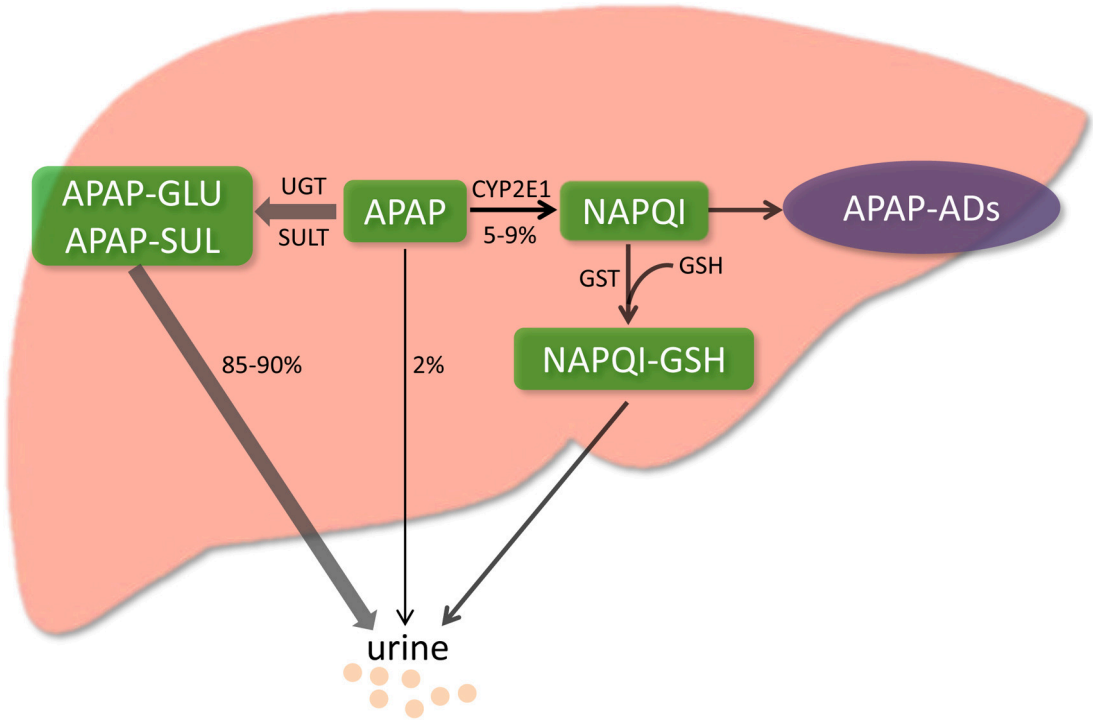

Figure 4.1.2.1 APAP metabolism and mechanisms of hepatotoxicity.
In the liver, APAP is mainly metabolised by UDP-glucoronosyltransferase (UGT) and sulfotransferase (SULT) into non-toxic glucuronide (APAP-GLU) and sulfate (APAP-SUL) prior to excretion in the urine. A small proportion (around 2%) is excreted as APAP unchanged in the urine. Around 5-9% of APAP is metabolised by cytochrome P450 enzymes, mainly CYP 2E1 into a highly reactive metabolite, called N-acetyl-p-benzoquinone imine (NAPQI). NAPQI is quickly metabolised into into a non-toxic metabolite by conjugation with glutathione (GSH), catalysed by glutathione-s-transferase (GST). In APAP overdose setting, excessive NAPQI is generated, leading to saturation of GST. NAPQI then binds to cellular protein, forming APAP protein adducts (APAP-ADs) and leading to hepatic cell death and injury. Illustration taken from [104].

Following APAP overdose, an antidote, called acetylcysteine, can prevent liver injury when administered within a few hours of overdose. The decision to treat APAP overdose patients with acetylcysteine is currently based on the levels of serum biomarker ALT and APAP concentration. However, current markers of APAP overdose are lacking in sensitivity and specificity when measured soon after APAP overdose, such as at initial presentation to hospital [102]. Pre-

clinical and clinical studies of APAP overdose have identified new potential biomarkers that may provide enhance specificity, such as miR-122, or inform mechanistic events, such as keratin-18 for necrosis, GLDH for mitochondrial dysfunction, acetyl-high mobility group box 1 protein (acetyl-HMGB1) for inflammation, as well as macrophage colony stimulating factor 1 (CSF-1) for tissue regeneration [103]. The utility of the analysis of cfDNA as a biomarker of APAP overdose has not previously been explored.

Similar to human, APAP overdose in mice may also cause acute liver injury, making mice a very suitable experimental animals for studying metabolic function and response following APAP overdose. Further, methodology for inducing APAP overdose in mice is well studied [105, 106]. With the consideration that there is a clinical need for biomarker of APAP overdose and practical ease of doing the experiment in mouse models, in this chapter, APAP overdose in hepatocyte-specific reporter mice was performed as a validation of results in Chapter 3.

4.2 Induction of APAP overdose in C57BL/6 mice

Prior to induction of APAP overdose in hepatocyte-specific reporter mice, a preliminary experiment in C57BL/6 mice was performed to investigate the timepoint following APAP dosing whereby cfDNA levels in plasma is highest. C57BL/6 mice was injected with standard dose APAP (350mg/kg of mouse body weight). This dose allows for non-lethal liver injury in C57BL/6 mice. Mice were culled and samples were collected at 8, 24 or 48 following APAP dosing. Analysis of cfDNA levels showed that the highest increase of cfDNA

concentration (~100-fold) was observed in APAP-dosed mice culled at 8 hour compared to mice receiving saline injection (**Figure 4.2.1a**). Analysis of protein biomarkers of liver function test ALT and AST showed that liver injury is apparent at each timepoint following APAP overdose, with the highest protein concentration at 8 hour and lowest at 24 hour (**Figure 4.2.1b-c**). No changes in the levels of liver biomarker ALB and Bilirubin, indicating the expected absence of liver injury at this APAP dose (**Figure 4.2.1d-e**). Histological liver sections at each timepoint after APAP dosing (**Figure 4.2.1f**) confirmed the presence of acute zone 3 coagulative necrosis at each timepoint starting at 8 hour following APAP dosing. The most severe injury was observed at 24 hour following APAP dosing. At 48 hour following APAP dosing, liver started regenerating. The timepoint 8 hour following APAP dosing was selected to be used in hepatocyte-specific reporter mice considering the peak levels of cfDNA.

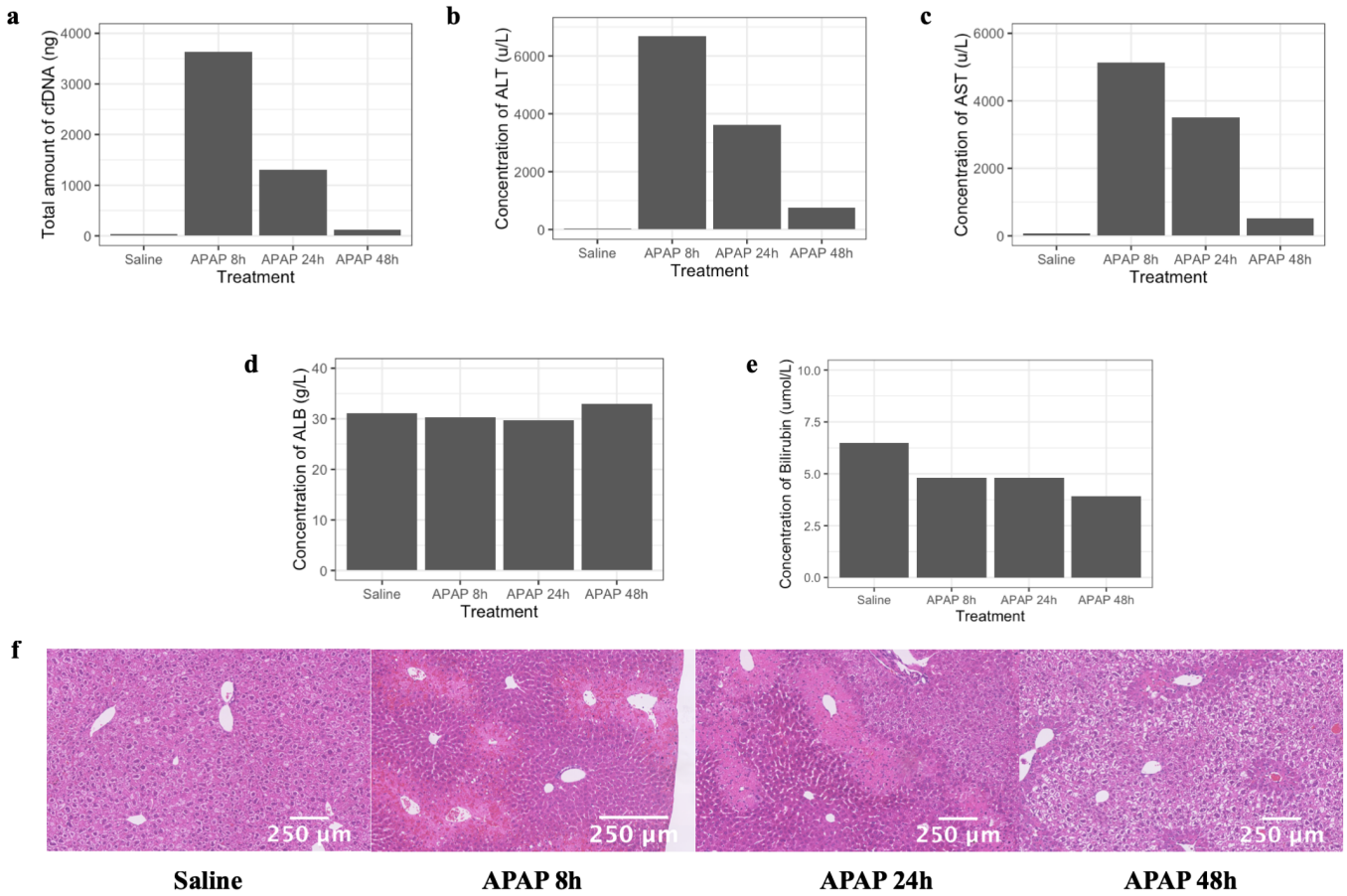

Figure 4.2.1 The effect of APAP on biomarkers at different timepoints in C57BL/6 mice.
The concentration of cfDNA (**a**), concentration of ALT (**b**) and AST (**c**) peaked at 8 hours following APAP dosing and decreased after 24 and 48 hours. The levels of ALB (**d**) and bilirubin (**e**) were unaffected by APAP dosing, indicating the absence of overt liver failure. Histological analysis (**f**) showed that liver damage (acute zone 3 coagulative necrosis typical of APAP hepatotoxicity) was most severe at 24-hour following APAP compared to other time points.

4.3 Analysis of Cre recombination in cfDNA of hepatocyte-specific reporter mice following APAP dosing

Similar to APAP overdose in C57BL/6 mice, analysis of total plasma cfDNA demonstrated an increase of ~100-fold in APAP-dosed hepatocyte-specific reporter mice compared to mice receiving saline injection (**Figure 4.3.1a**). Analysis of liver biomarkers, and histological analysis of mouse liver tissues confirmed the presence of liver injury (**Figure 4.3.1b-f**). Furthermore, quantification of liver-specific cfDNA using 1lox/2lox ddPCR assay at 8 hours post APAP dosing in hepatocyte-specific reporter mice shows hepatocyte contribution to cfDNA increased to 77.9%, compared to 4.2% in non-treated healthy mice. This increase was not caused by IP injection (**Figure4.3.1g**).

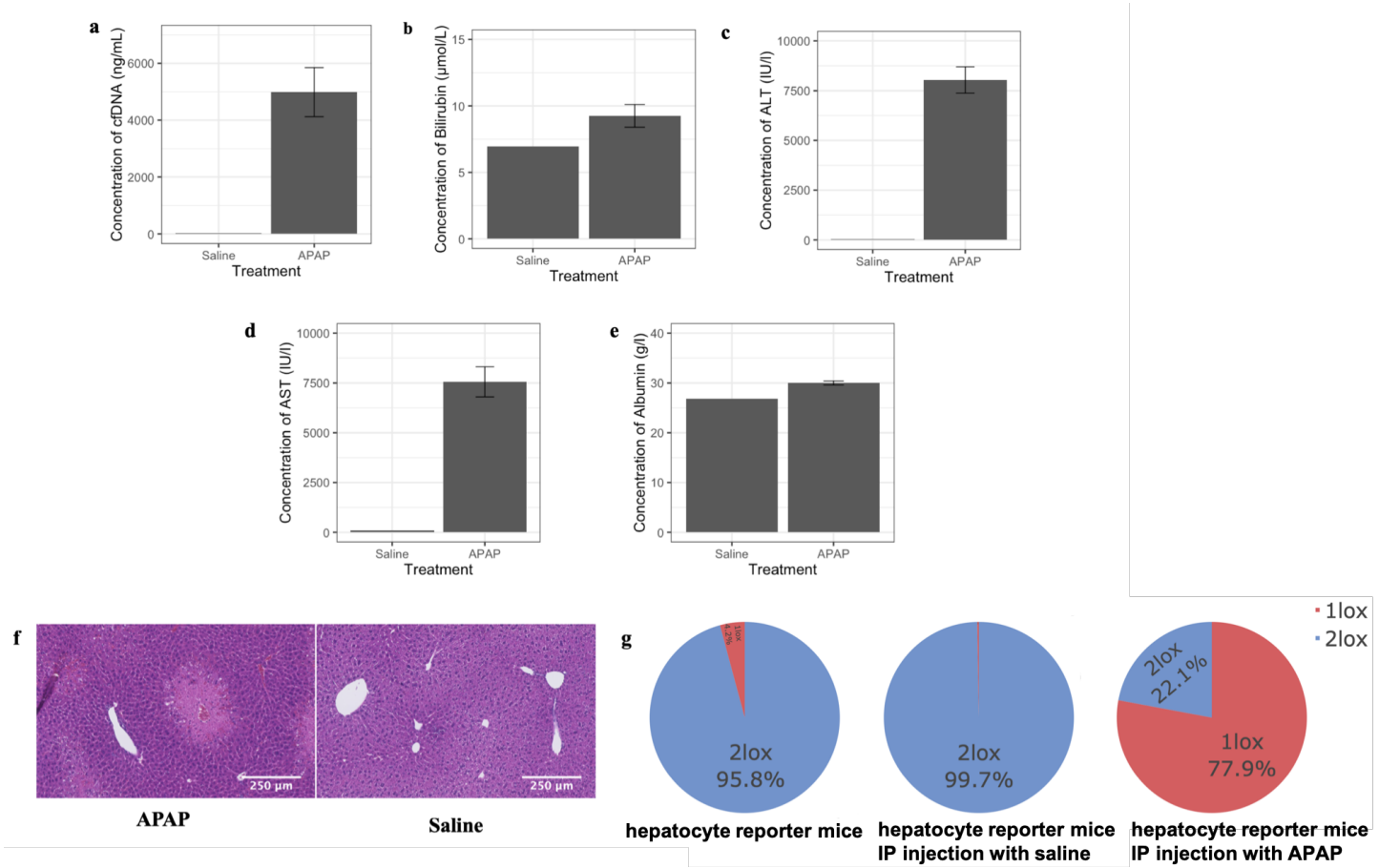

Figure 4.3.1 The tissue origins of cfDNA following APAP overdose in mouse models.
Barplots of mean concentration of blood biomarkers from APAP-injected hepatocyte-specific reporter mice (n = 4) and saline-treated negative control (n = 10, pooled) at 8 hour following APAP dosing: (**a**) total cfDNA, (**b**) bilirubin, (**c**) AST, (**d**) ALT, (**e**) ALB. Error bars showed standard error. (**f**) H&E staining of liver tissues from an APAP-injected hepatocyte-specific reporter mice vs saline-treated negative control showing acute zone 3 coagulative necrosis typical of APAP hepatotoxicity. (**g**) Comparison of hepatocyte cell contribution in healthy mice vs post APAP overdose (n = 4).

4.4 Analysis of cfDNA in APAP overdose patients

A liver-specific methylation ddPCR assay [84] was validated to analyse cfDNA of APAP overdose patients. The assay amplified a region of chromosome 8 in the human genome where around 60% of methylation density/percentage was observed in liver as opposed to other tissues, such as blood cells, bladder, oesophagus, heart, lung, pancreas and small intestine (**Figure 4.4.1a**). Validation of this assay was performed on gDNA from liver and white blood cells. The probe that binds to methylated CpG sites originating from liver showed high specificity to liver (**Figure 4.4.1b**). The probe that binds to unmethylated CpG sites originating from other tissues, showed non-specific amplification in liver (96 positive droplets out of 14266, **Figure 4.4.1c**).

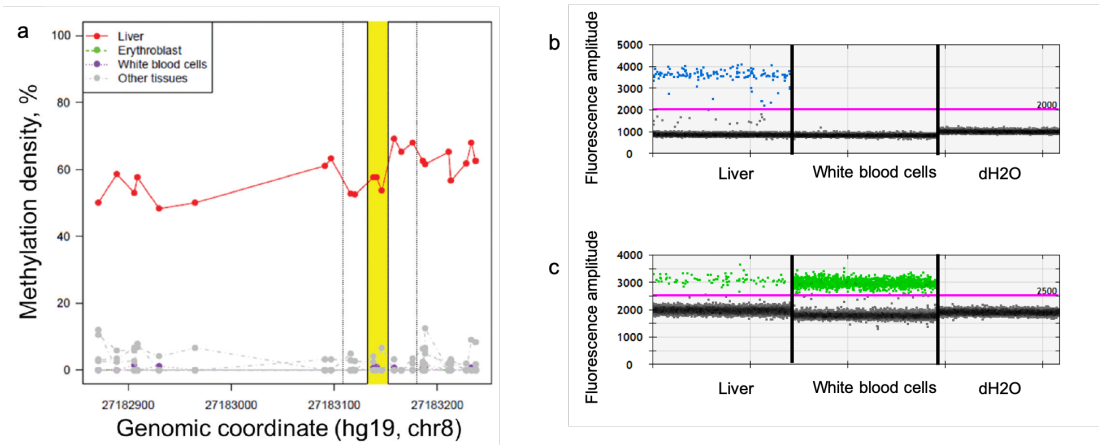

Figure 4.4.1 Specificity and validation of liver-specific methylation ddPCR assay. (**a**) Genomic coordinate of liver-specific methylation marker and its specificity compared to blood and other tissues (bladder, oesophagus, heart, lung, pancreas and small intestine). Illustration taken from [84]. Validation of liver-specific methylation assay on gDNA from liver and white blood cells. (**b**) Liver-specific probes bind to methylated CpG sites in liver and not in white blood cells. (**c**) Non-liver probes bind to white blood cells, with few non-specific amplicons in liver.

To demonstrate potential applicability of these findings to APAP overdose in humans, the total amount of cfDNA and protein liver function biomarkers were

measured in the serum of healthy volunteers (n = 11) and APAP overdose patients (n=8), a subset of which (n = 4) were not exhibiting clinically apparent liver injury based on serum ALT levels (**Appendix: Table B.1**). On average, the concentration of cfDNA in the APAP overdose patients increased by ~126-fold (**Figure 4.4.2a**). In patients with clinically apparent liver injury (high ALT), total cfDNA increased by ~234-fold, whereas patients without clinically apparent liver injury (normal ALT) demonstrated mean total cfDNA increase of ~18-fold compared to healthy volunteers (**Figure 4.4.2a**). The increase in total cfDNA following APAP overdose in patients was consistent with the increase in cfDNA following APAP in C57BL/6 and hepatocyte-specific reporter mice, indicating the potential of analysis of cfDNA as a clinical biomarker in APAP overdose.

To further assess the utility of cfDNA as a clinical biomarker, serum cfDNA concentration was compared with other conventional biomarkers of liver function. Total cfDNA was higher in all patients with APAP overdose than in healthy volunteers, although two samples, one from an overdose patient with normal ALT (cfDNA = 192.87) and one healthy volunteer (cfDNA = 188.17) showed similar cfDNA concentrations (**Figure 4.4.2b**). Total concentration of cfDNA also correlated with levels of GLDH, a novel APAP overdose biomarker that measures mitochondrial damage (Spearman correlation 0.85, p-value <0.01) with the (**Figure 4.4.1c**).

Specific analysis of liver contributions to the cfDNA pool, measured by ddPCR, showed the percentage of liver-specific cfDNA fragments increased by ~175-fold in patients with clinically apparent liver injury compared to healthy

volunteers, but no increase in liver-specific cfDNA was observed in overdose patients without clinically apparent liver injury (normal ALT) (**Figure 4.4.2d, e**). Although GLDH levels were increased in APAP overdose patient with normal ALT, analysis of liver-specific cfDNA indicated that the cfDNA did not originate from liver tissue (**Figure 4.4.2**).

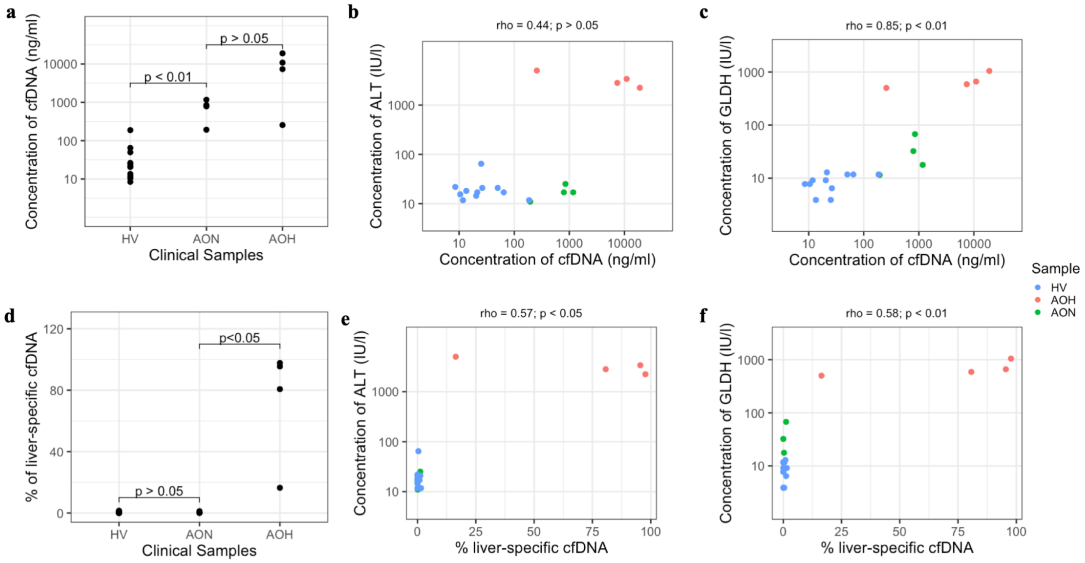

Figure 4.4.2 Analyses of cfDNA and other liver biomarkers in clinical samples. Comparison of (**a**) total concentration of cfDNA between healthy volunteers (HV), APAP overdose patients with normal ALT (AON) and high ALT (AOH) and (**b**) comparison of serum ALT and total concentrations of cfDNA, (**c**) comparison of serum GLDH and total concentrations of cfDNA, (**d**) comparison of percentage of liver-specific cfDNA between samples, (**e**) comparison of serum ALT and percentage of liver-specific cfDNA, (**f**) comparison of serum GLDH and percentage of liver-specific cfDNA. Statistical significance for panel (**a**) and (**d**) were obtained from Mann-Whitney U test and for panel (**b**), (**c**), (**e**), (**f**) from Spearman's rank correlation coefficient.

4.5 Discussion

In this chapter, the utility of tissue-specific reporter mice to measure the tissue origins of cfDNA in a pathological scenario was successfully demonstrated. Following APAP overdose, an increase of both total concentration of cfDNA (around 100-fold) and hepatocyte-derived cfDNA fragments (around 20-fold) were observed. This showed definitively that tissues with high turnover and

more cellular death contributed more than tissue with lower turnover and less cellular death. This also indicated that cfDNA is released by cellular death mechanisms.

The utilisation of mouse models in this study enabled control over variables that affects the outcome of the experiment compared to study in human, such as age, sex, and dosage of APAP. APAP overdose was induced in male hepatocyte-specific reporter mice aged 10-12 week old. The induction of tissue injury using APAP was selected in this study because mechanisms of hepatotoxicity from APAP overdose is well described [104]. The dosage of APAP in this study was selected to be non-lethal for C57BL/6J mice. Further, only male mice aged 10-12 week old were used to control for known variations of response to APAP in male and female mice, with female mice developing less liver injury than male mice at the same dosage of APAP [106]. The selection of defined sex and age in this experiment also allows for direct comparison of tissue contributions from hepatocyte in healthy hepatocyte-specific reporter mice in Chapter 3.

Despite successful demonstration of tissue-specific reporter mice as a model to measure the tissue origins of cfDNA, this experiment has limitations. Measurement of APAP overdose in mice was done using 4 technical replicates at a single, non-lethal dose and culling was performed only at one timepoint following APAP overdose. Further experiments with more technical replicates and a combination of APAP dose at different timepoints may provide more information on the effect of different dosages of APAP over time. Using this model, the increase of tissue contribution can only be investigated for

hepatocytes. Induction of APAP overdose using other tissue-specific reporter mice may be beneficial to show toxicity of APAP outside of the liver because it is an area of research that is not well studies to date [107].

Measurement of the contribution of hepatocytes/liver in hepatocyte-specific reporter mice may also be performed using hepatocyte-specific methylation biomarker as an alternative to the floxed *mT/mG* assay. This method would require a comparison of many tissues/cell types in mice and a deconvolution step to screen for a region with specific CpG dinucleotide methylation status when compared to other tissues. Prior to utilisation in cfDNA, a methylation-specific ddPCR assay would need to be designed and validated on bisulphite-converted mouse gDNA. Compared to direct measurement of the floxed *mT/mG* alleles, this approach would involve more steps and could introduce more errors. However, measurement using this methylation assay may validate measurements of the floxed *mT/mG* alleles. Further, measurement using methylation ddPCR assay may be more applicable to human studies that uses similar methylation approach.

The applicability of findings in mice to human studies was also demonstrated. Similar to findings in mice, total concentration of cfDNA was increased in 8 APAP overdose patients. Previous studies have suggested the utility of cfDNA as a clinical biomarker of cell death in a range of clinical conditions [32, 108], this study was the first demonstration of the clinical utility of total cfDNA as a biomarker for APAP overdose in human. Analysis of total concentration of cfDNA showed that total concentration of cfDNA robustly distinguished APAP overdose patients from healthy volunteers, suggesting that total cfDNA

concentration may reflect exposure to APAP. In overdose patients with raised ALT, the increase in cfDNA concentration was greater than in those with normal ALT, although all overdose patients had a higher cfDNA concentration than healthy volunteers. However, whilst there was also a clear increase in the proportion of cfDNA that originated from liver in overdose patients with raised ALT concentrations, the proportion of liver-specific cfDNA was not increased in overdose patients with normal ALT. This contrasts with the observation of increased GLDH in all overdose patients, both those with raised and normal ALT.

Previous studies have proposed GLDH as a marker of liver damage in APAP overdose [109], but these data suggest that in APAP patients with normal ALT, the rise in total cfDNA and GLDH was not liver-derived and may have been due to damage of tissues other than the liver. Given that serum GLDH is a measure of mitochondrial damage [110] and that GLDH is expressed heterogeneously across mammalian tissues [111], the increase in GLDH and cfDNA in patients without clinically overt liver injury may reflect release of cfDNA from other tissues, for which further pre-clinical and clinical work is indicated to establish the source. To establish the source of cfDNA following APAP overdose, analysis in blood-specific and other reporter lines, such as kidney, may provide useful information. Similarly, the contribution of different tissues in APAP overdose patients may be analysed using various tissue-specific methylation assays.

Chapter 5 Screening for somatic mutations as candidate liquid biopsy biomarkers in the colonic adenocarcinoma of the Pirc rat

5.1 Introduction

Successful implementation of cancer liquid biopsy requires not only a deep understanding of the analyte biomolecule, such as cfDNA, but also requires promising tumour-specific alterations to be analysed. Unlike previous result chapters, in Chapter 5, a genetically altered rat model of colorectal cancer (CRC) is screened for potential candidate biomarkers of liquid biopsy. This chapter serves as a preliminary work for applying the analysis of cfDNA in the Pirc rat. Candidate biomarkers found in this chapter may be used to further develop the Pirc rat as an animal model to study and develop non-invasive detection methods of colorectal cancer.

5.1.1 Familial adenomatous polyposis

CRC is one of the most common types of cancer in the world. Approximately 35% of all CRC cases are associated with a genetic predisposition. Familial adenomatous polyposis (FAP) is an inherited form of CRC, which constitutes 1% of all CRC cases. FAP patients develop adenomas at an unusually young age. One or more of these adenomas become malignant with almost 100% penetrance when left untreated [112].

The majority of FAP patients have mutations in the adenomatous polyposis coli (*APC*) gene. The *APC* gene is a tumour suppressor gene that plays a central role in the Wnt signalling pathway, specifically in the

degradation of β-catenin in the cytoplasm. Mutation in the *APC* gene can affect the suppression β-catenin and lead to constitutive activation of several genes and oncogenes that control cellular growth and division. This results in cellular overgrowth and formation of adenomas [112]. A proportion of FAP patients carry a heterozygous germline mutation in the *APC* gene. A somatic mutation in the *APC* gene through allelic loss or truncating deletion serves as a "second hit" to the gene and presumes to initiate the growth of these tumours [113]. Another proportion of FAP patients have no mutation in the *APC* gene. Currently, identification of novel genes associated with FAP is ongoing [112].

5.1.2 The Pirc Rat: a model for familial adenomatous polyposis

The Pirc rat ($Apc^{Pirc/+}$) is a rodent model of FAP. Similar to a majority of FAP patients, the Pirc rat has a heterozygous germline mutation at the *Apc* gene, specifically at nucleotide 3409. The mutation changed the nucleotide from adenine to thymine, causing a change in amino acid sequence from lysine (AAG) to stop signal (TAG) during translation, This mutation causes the rat to develop colonic tumours as early as 45 days of age [114].

The Pirc rat was generated through the injection of germline mutagen N-ethyl-nitrosourea (ENU) on male inbred F344/NTac rats. Progeny of these rats were screened for truncating mutation in the *Apc* gene [114]. The Pirc rats used in this project were ACI/NHsd congenic strain. This was achieved by crossing of the Pirc rat with F344/NTac background to inbred ACI/NHsd background, and backcrossing to inbred ACI/NHsd rats for at least 17

generations to remove residual heterozygosity (personal communication with Dr. James Amos-Landgraf).

Colonic tumours in the Pirc rat (*Apc*$^{Pirc/+}$) develop through two main mechanisms: LOH of the short arm of chromosome 18 without changes in copy number, and monoallelic expression of mutant *Apc* allele in tumours that maintain heterozygosity [69, 115]. Outside the *Apc* region, the genomic landscape in colonic tumours of the Pirc rat has not previously been described. Additionally, characterisation of LOH in the colonic tumour of the Pirc rat at a single nucleotide resolution has not been performed. WGS of colonic tumours from the Pirc rat may potentially reveal novel genes associated with FAP and provide characterisation of LOH at single nucleotide resolution. In this chapter, WGS was performed to characterise the genomic landscape of CRC in the Pirc rat and screen for candidate biomarkers for liquid biopsy.

5.2 Histological validation of colonic tissue identity

Identity of colonic tissues in this study were validated by histological analysis of H&E-stained tissue sections. A total of 46 colonic tumours from 21 rats were identified. A representative tissue section of normal colon tissue and colonic tumour are shown in **Figure 5.2.1**. Colonic tumour tissues were distinguished from normal colon tissues by the presence of irregularly-shaped intestinal glands in the mucosa of tumour tissues.

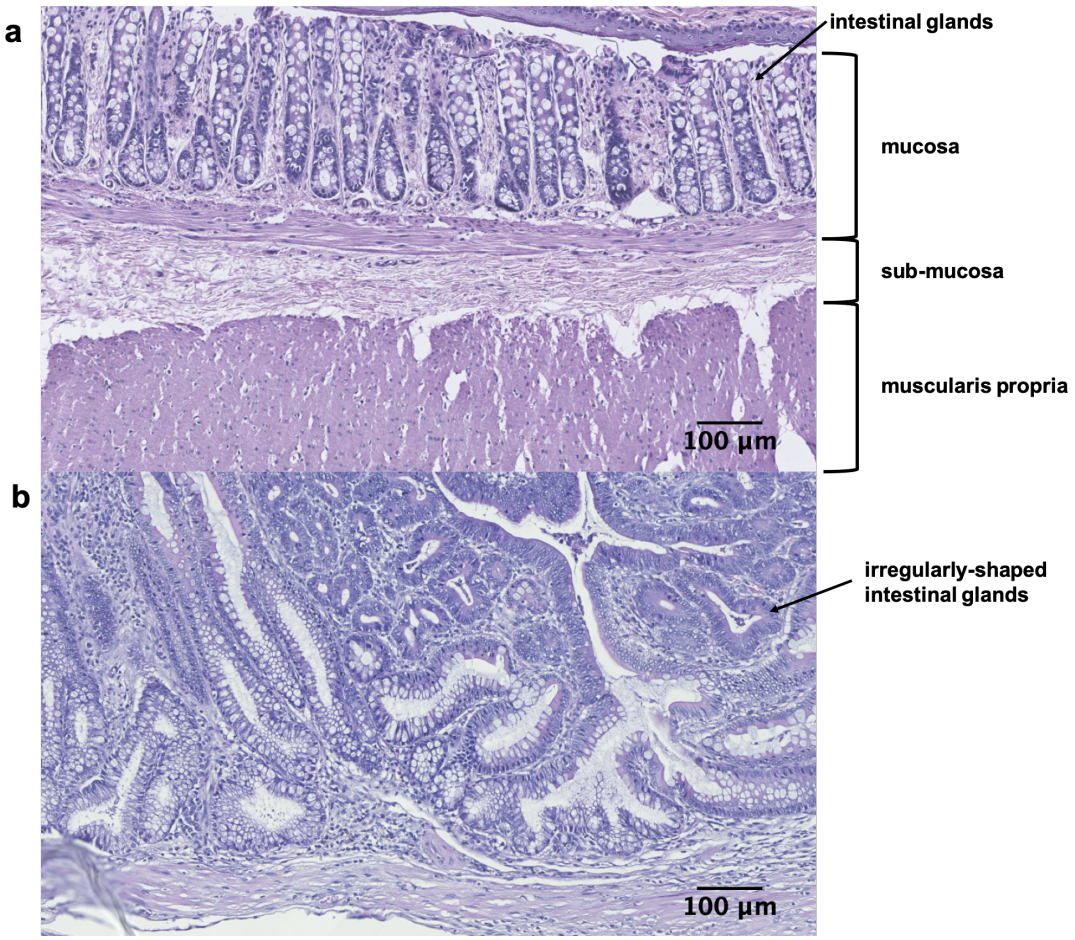

Figure 5.2.1 H&E-stained normal colon and colonic tumour tissues from the Pirc rat. (a) a section of normal colon tissue, showing three layers of colon: mucosa, sub-mucosa and muscularis propria. Intestinal glands formed by epithelial cells are found in the mucosa layer. (b) a section of colonic tumour tissue, with irregularly-shaped intestinal gland, a characteristic of neoplastic colon tissue.

5.3 Whole genome sequence generation and variant calling

WGS of paired liver-colonic tumour DNA samples from 2 rats was performed to identify somatic alterations in the colonic tumour of the ACI-Pirc ($Apc^{Pirc/+}$) rat. WGS was also performed on liver DNA from ACI-WT ($Apc^{+/+}$) rat. These sequences were analysed alongside F344N/Crl sequence available from a public database [85] because the congenic region of the rats was heterozygous for ACI and F344N genome. Table 1 shows the summary of the

WGS coverage, and the total number of SNVs and small indels relative to the Brown Norway genome reference (RGSC Rnor_6.0). Average coverage for tumour samples were above 60x, and germline samples were above 30x. Across all samples, more than 70% of the genome were covered at the expected coverage. Quality of sequences from Illumina HiSeqX is available in Appendix A (**Figure A.3-5**). High quality sequences were obtained with base quality score more than 30.

Comparison of the percentage of heterozygous variants in the germline DNA of the Pirc rat with an inbred rat strain F344N/Crl (**Table 5.3.1**) showed that germline Pirc rat DNA had a higher percentage of heterozygous SNVs and indel than F344N/Crl. The difference in the heterozygous variant percentages was likely due, in part, to the presence of congenic region in the Pirc rat DNA. Visualisation of heterozygous SNVs across chromosomes (**Figure 5.3.1 – Figure 5.3.5**) confirmed the presence of a large region with heterozygous SNVs in chromosome 18, and patches of heterozygous SNVs across other regions of the genome. Even after backcrossing for at least 17 generations, the Pirc rats used in this study showed almost 20% of heterozygosity across the genome. This rate of heterozygosity is higher than what one would expect for inbred lines. This high heterozygosity across the genome may indicate breeding errors made during backcrossing or lack of complete inbreeding of the parental strains. Fortuitously, this high level of heterozygosity enabled screening for LOH to be performed not only in the congenic region (at chromosome 18) but also across the genome.

Table 5.3.1 Summary of whole genome sequence coverage and SNVs and small indels relative to the Brown Norway reference genome.

Sample	Gb of bases mapped	Average Coverage	% of bases covered at expected coverage	SNV		Small Indels		Percentage of Heterozygous Variant (%)	
				Hom	Het	Hom	Het	SNV	Indels
12WTLIVER	106.79	39.12	77.7	3381597	698986	1344503	207911	17.13	13.39
9PIRCLIVER	120.6	44.18	84.6	3347211	782037	1331612	213231	18.94	13.80
13PIRCLIVER	101.76	37.28	74.8	3313410	779650	1310670	232314	19.05	15.06
9PIRCADENOCARCINOMA	226.32	82.91	84.7	3614367	845647	1759343	331725	18.96	15.86
13PIRCADENOCARCINOMA	216.2	79.2	83.2	3621487	844014	1775879	348696	18.90	16.41
F344N/Crl	73.72	27	72.5	2409365	271273	748454	60915	10.12	7.53

Sample description: 12WTLIVER was liver DNA from rat no. 12 ACI-WT (*Apc*$^{+/+}$). 9PIRCLIVER and 13PIRCLIVER were liver DNA from rat no. 9 and 13 ACI-Pirc (*Apc*$^{Pirc/+}$) respectively. 9PIRCADENOCARCINOMA and 13PIRCADENOCARCINOMA were colonic tumour DNA from rat no. 9 and 13 ACI-Pirc (*Apc*$^{Pirc/+}$) respectively. F344N/Crl genome sequence was obtained from Atanur *et al*. (2013). Hom: homozygous; het: heterozygous.

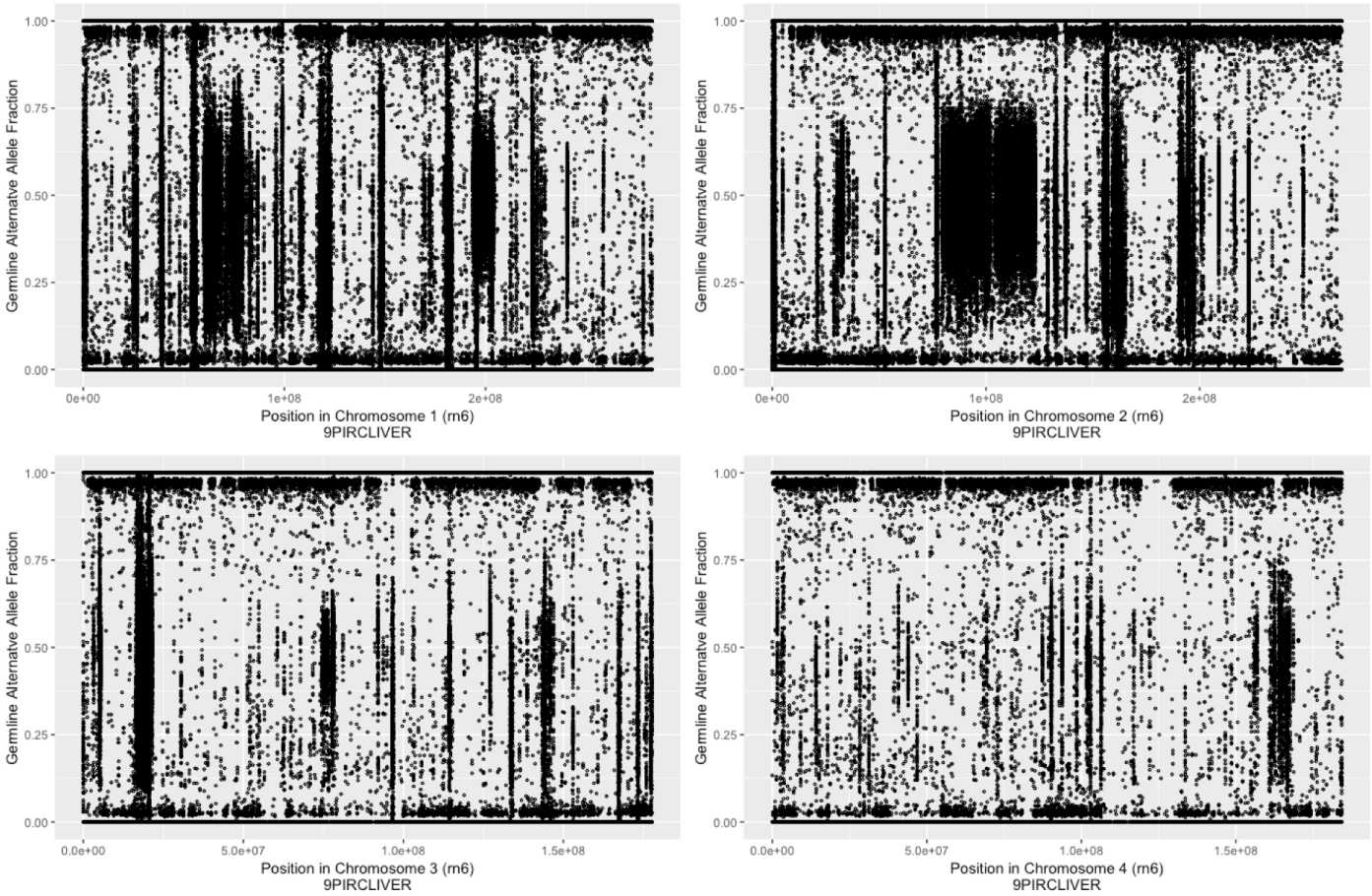

Figure 5.3.1 Pattern of heterozygosity in the Pirc rat genome from chromosome 1 to 4.
Each dot in the plot represent a SNV at a genomic position on the x-axis with alternate allele fraction displayed on the y-axis. The bolder dots appear on the plot means more SNVs. Homozygous SNVs are close to 0 or 1, heterozygous SNVs are between 0 and 1. Patches of heterozygous regions are observed across the chromosome.

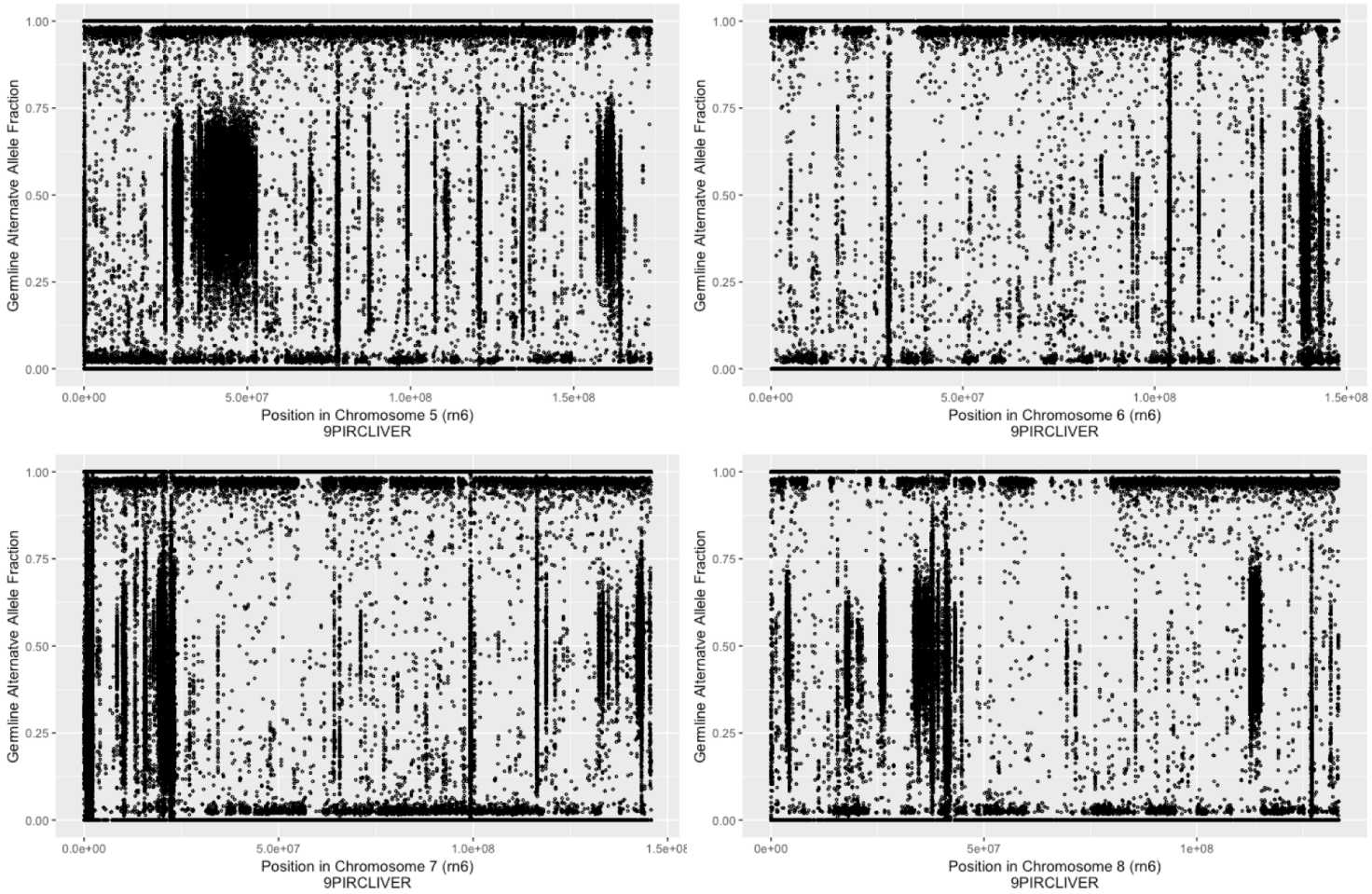

Figure 5.3.2 Pattern of heterozygosity in the Pirc rat genome from chromosome 5 to 8.
Each dot in the plot represent a SNV at a genomic position on the x-axis with alternate allele fraction displayed on the y-axis. The bolder dots appear on the plot means more SNVs. Homozygous SNVs are close to 0 or 1, heterozygous SNVs are between 0 and 1. Patches of heterozygous regions are observed across the chromosome.

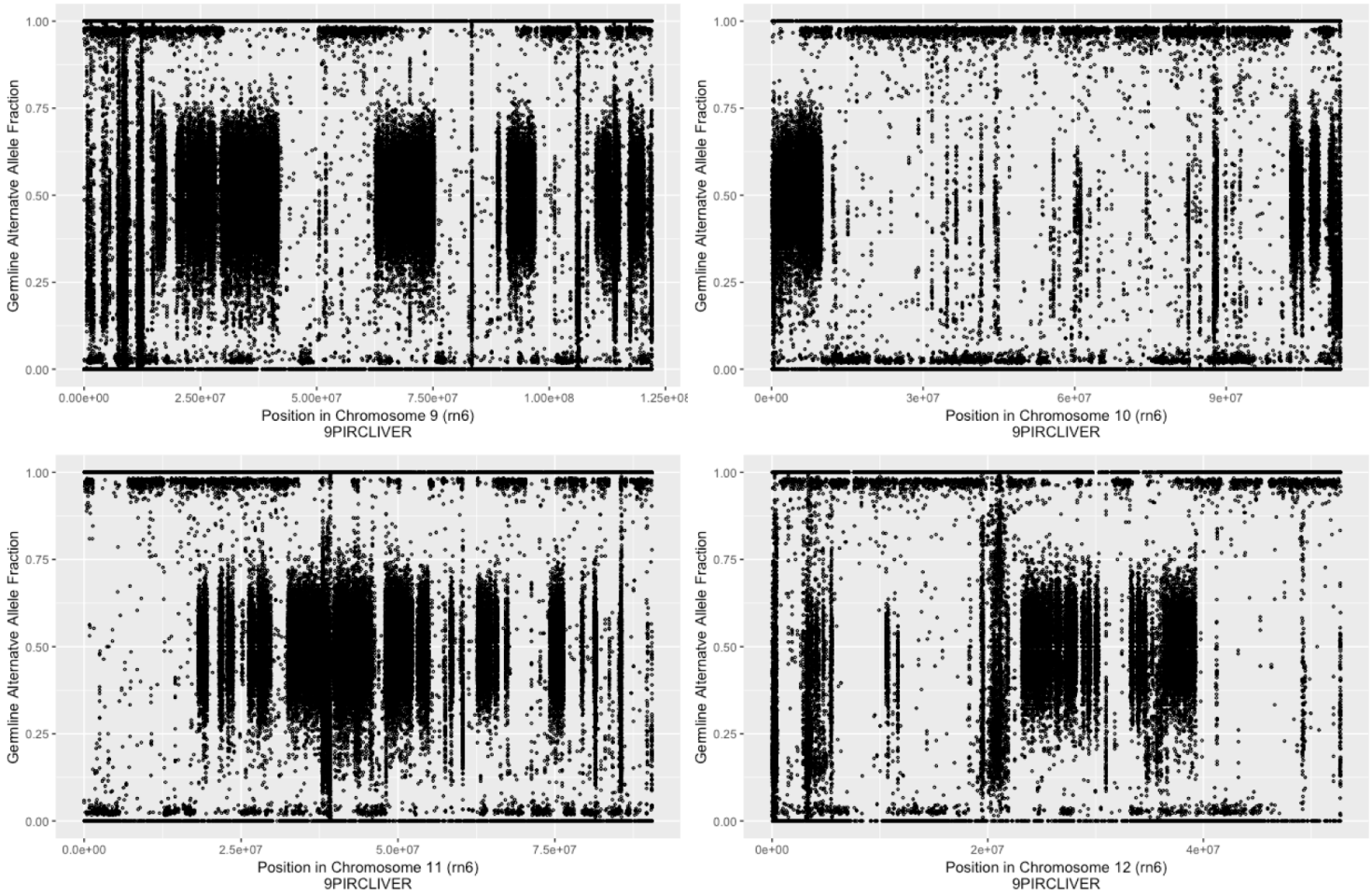

Figure 5.3.3 Pattern of heterozygosity in the Pirc rat genome from chromosome 9 to 12.
Each dot in the plot represent a SNV at a genomic position on the x-axis with alternate allele fraction displayed on the y-axis. The bolder dots appear on the plot means more SNVs. Homozygous SNVs are close to 0 or 1, heterozygous SNVs are between 0 and 1. Patches of heterozygous regions are observed across the chromosome.

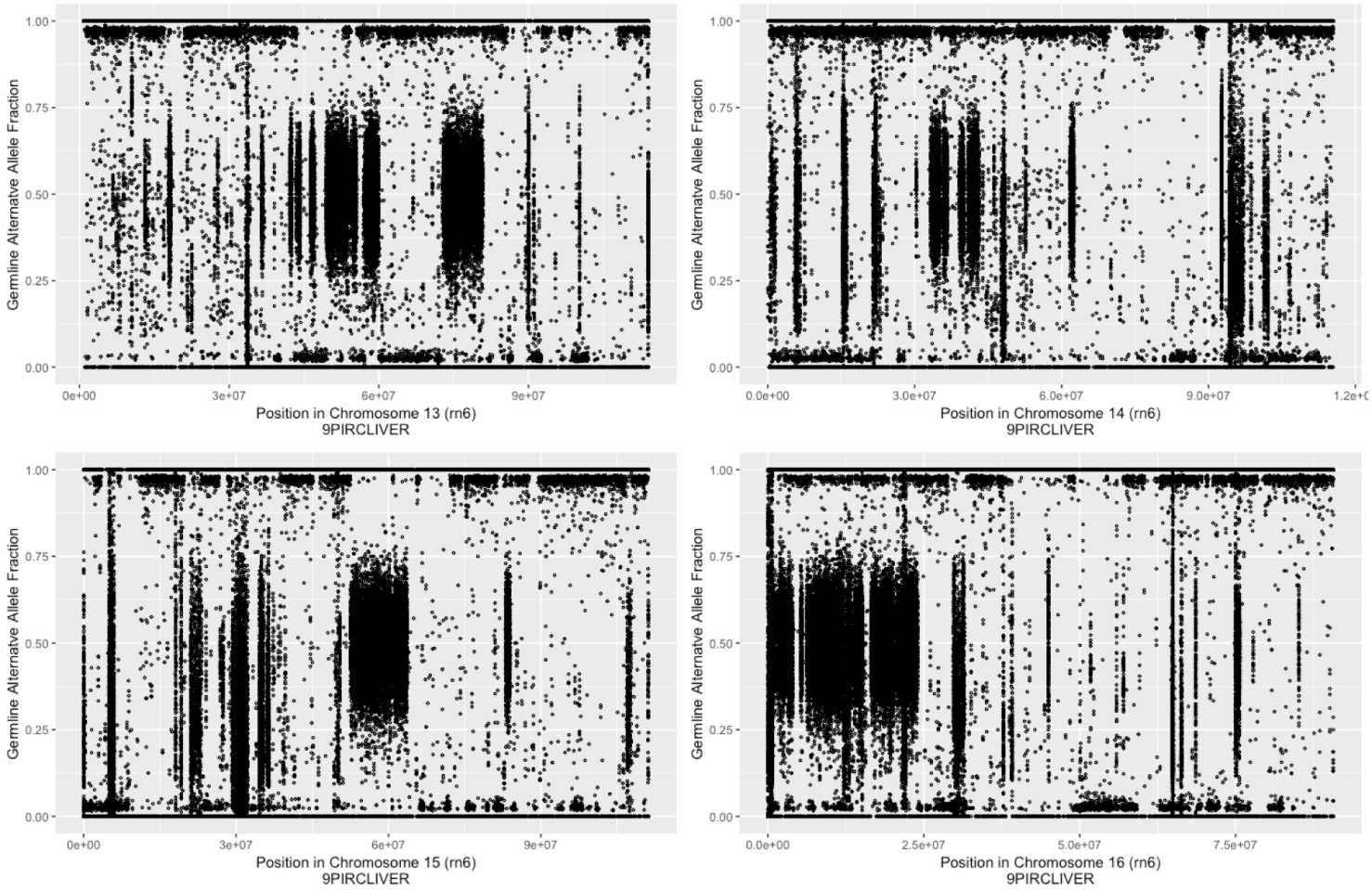

Figure 5.3.4 Pattern of heterozygosity in the Pirc rat genome from chromosome 13 to 16.
Each dot in the plot represent a SNV at a genomic position on the x-axis with alternate allele fraction displayed on the y-axis. The bolder dots appear on the plot means more SNVs. Homozygous SNVs are close to 0 or 1, heterozygous SNVs are between 0 and 1. Patches of heterozygous regions are observed across the chromosome.

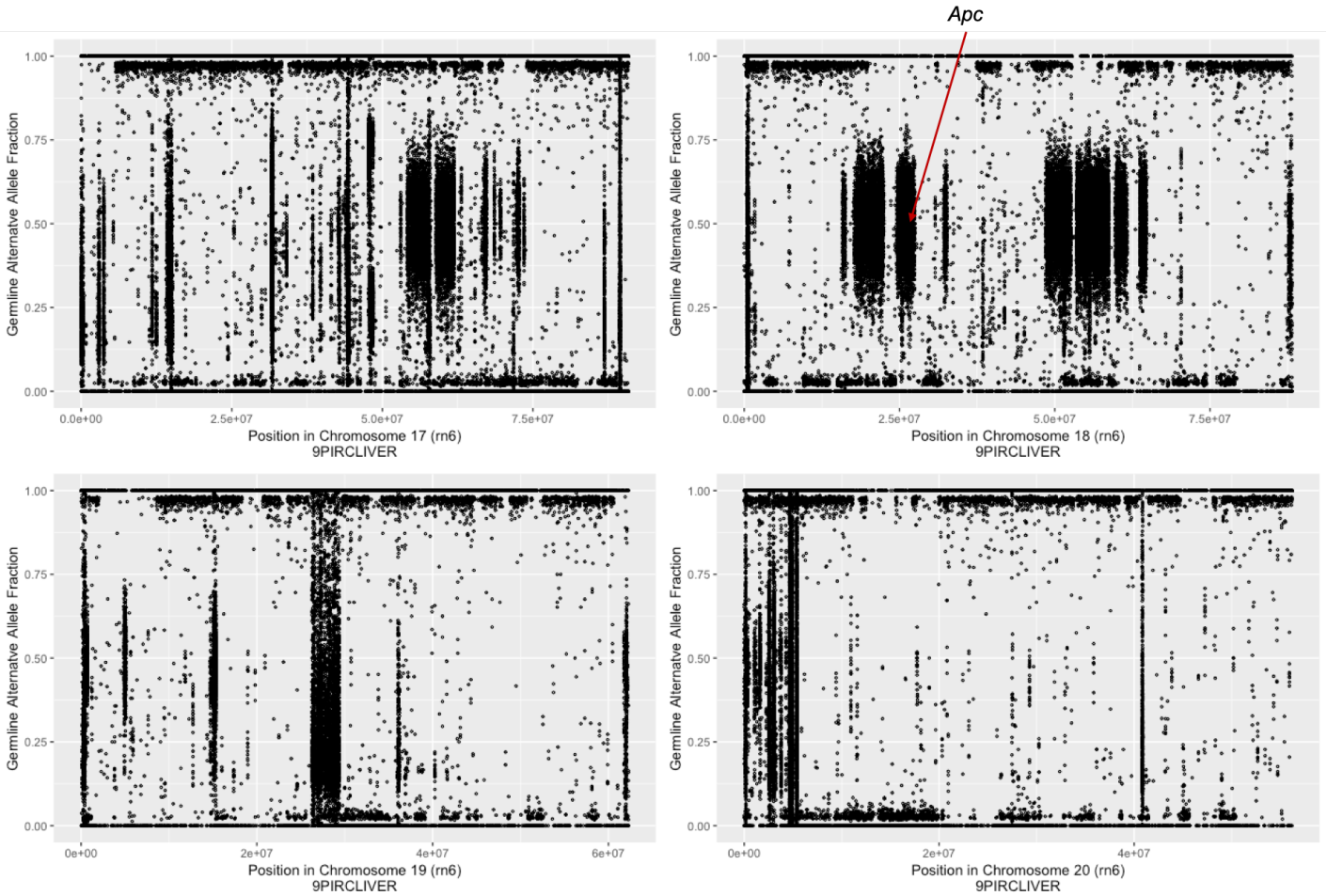

Figure 5.3.5 Pattern of heterozygosity in the Pirc rat genome from chromosome 17 to 20.
Each dot in the plot represent a SNV at a genomic position on the x-axis with alternate allele fraction displayed on the y-axis. The bolder dots appear on the plot means more SNVs. Homozygous SNVs are close to 0 or 1, heterozygous SNVs are between 0 and 1. Patches of heterozygous regions are observed across the chromosome. The location of the Apc gene is marked by a red arrow.

5.4 Screening, detection and validation of somatic LOH variants in the Pirc rat cancer genome

Comparison of allele balance of heterozygous SNVs were performed between paired liver and colonic tumour of each Pirc rat to identify regions with (LOH) in the cancer genome (**Figure 5.4.1** – **Figure 5.4.5**). Somatic LOH was found only around the Apc locus in chromosome 18 on the tumours from both Pirc rats and in no other chromosomal regions of the entire genome sequenced (**Figure 5.4.1** – **Figure 5.4.6**). A comparison of sequencing depth at the *Apc* Pirc locus (chr18:27100213) with the average genome coverage for each sample showed no substantial changes in sequencing depth (**Table 5.4.1**), indicating that the LOH occurred in the genome without copy number change and due to somatic recombination.

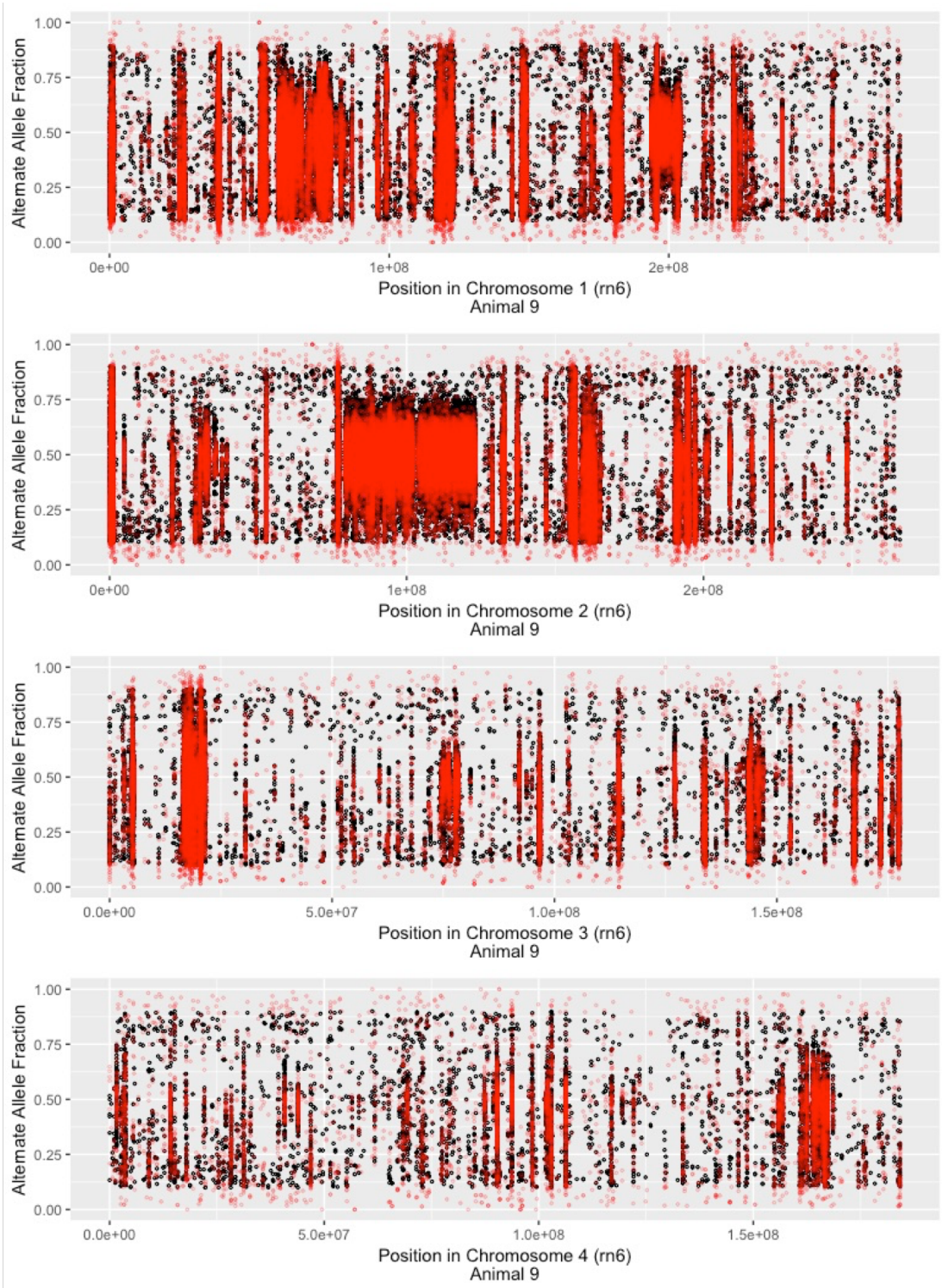

Figure 5.4.1 Screening for somatic LOH in a colonic tumour of the Pirc rat from chromosome 1 to 4.

The X-axis shows the position in chromosome 18, where the somatic LOH was observed. The Y-axis shows the alternate allele fraction of different SNVs. The alternate allele fraction of SNVs from germline DNA is depicted with black dots, and tumour DNA with red dots. The discontinuous SNVs density was due to the absence of heterozygous SNVs the loci.

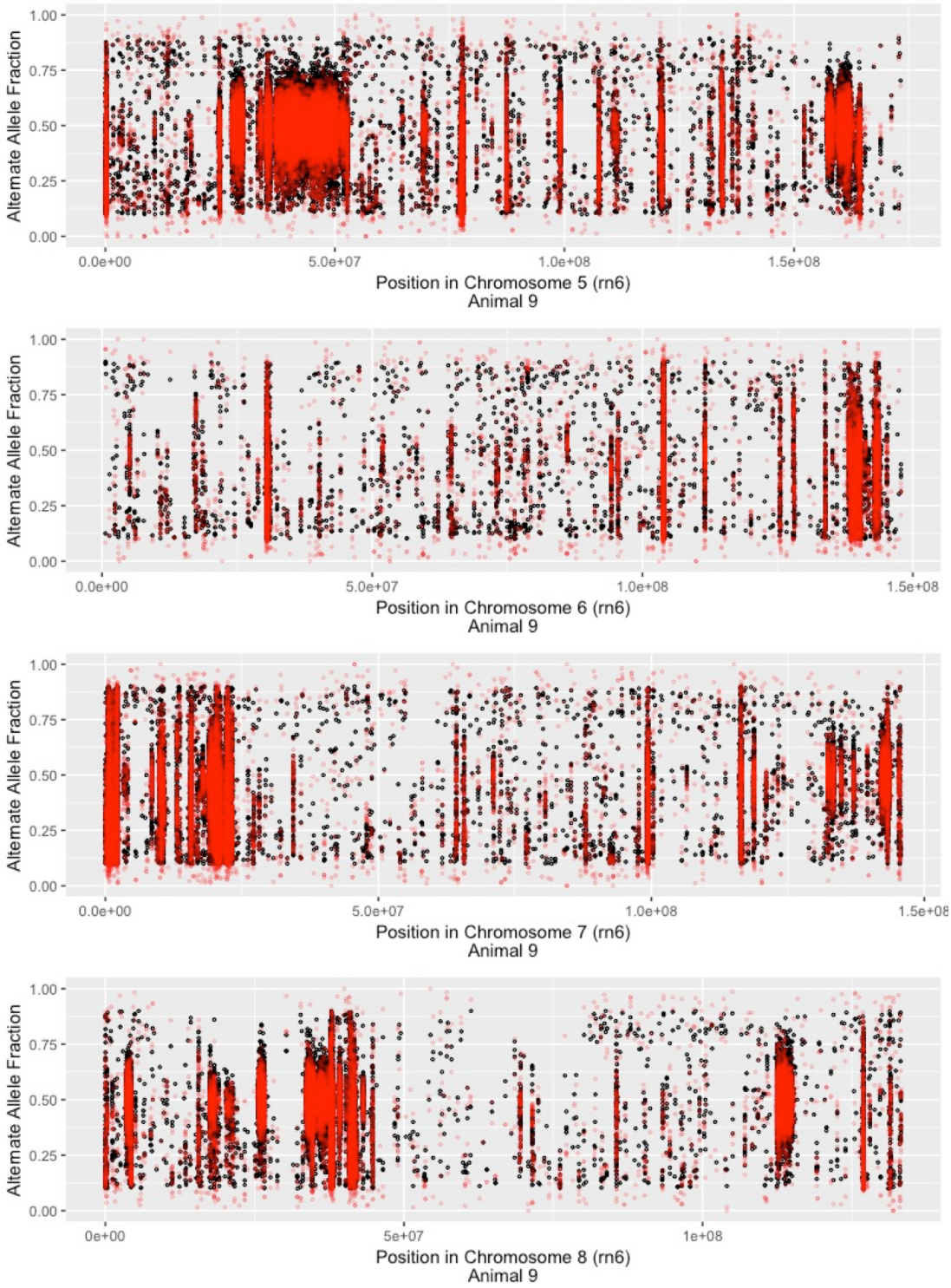

Figure 5.4.2 Screening for somatic LOH in a colonic tumour of the Pirc rat from chromosome 5 to 8.
The X-axis shows the position in the chromosome, where the somatic LOH was observed. The Y-axis shows the alternate allele fraction of different SNVs. The alternate allele fraction of SNVs from germline DNA is depicted with black dots, and tumour DNA with red dots. The discontinuous SNVs density was due to the absence of heterozygous SNVs the loci.

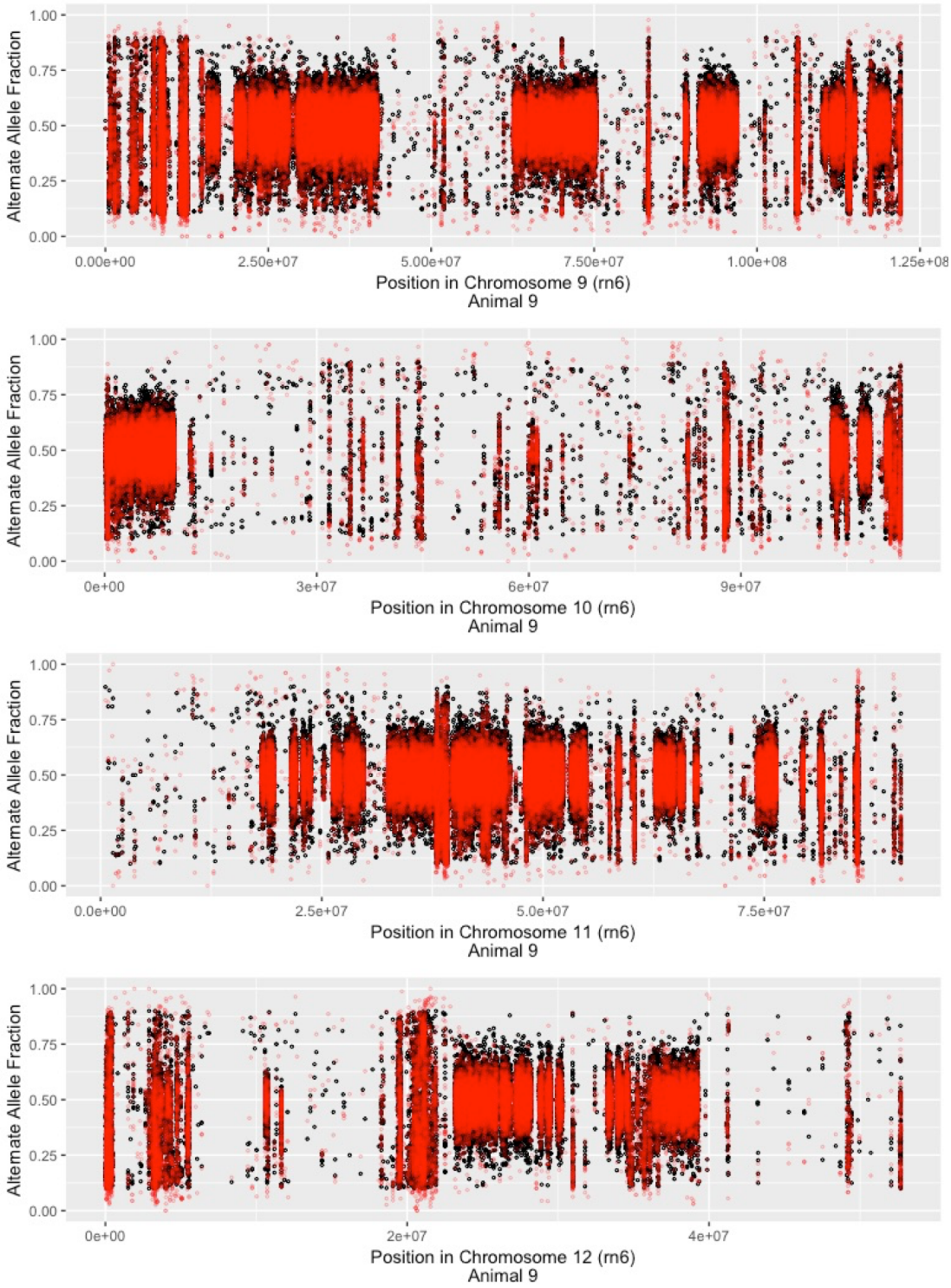

Figure 5.4.3 Screening for somatic LOH in a colonic tumour of the Pirc rat from chromosome 9 to 12.
The X-axis shows the position in the chromosome, where the somatic LOH was observed. The Y-axis shows the alternate allele fraction of different SNVs. The alternate allele fraction of SNVs from germline DNA is depicted with black dots, and tumour DNA with red dots. The discontinuous SNVs density was due to the absence of heterozygous SNVs the loci.

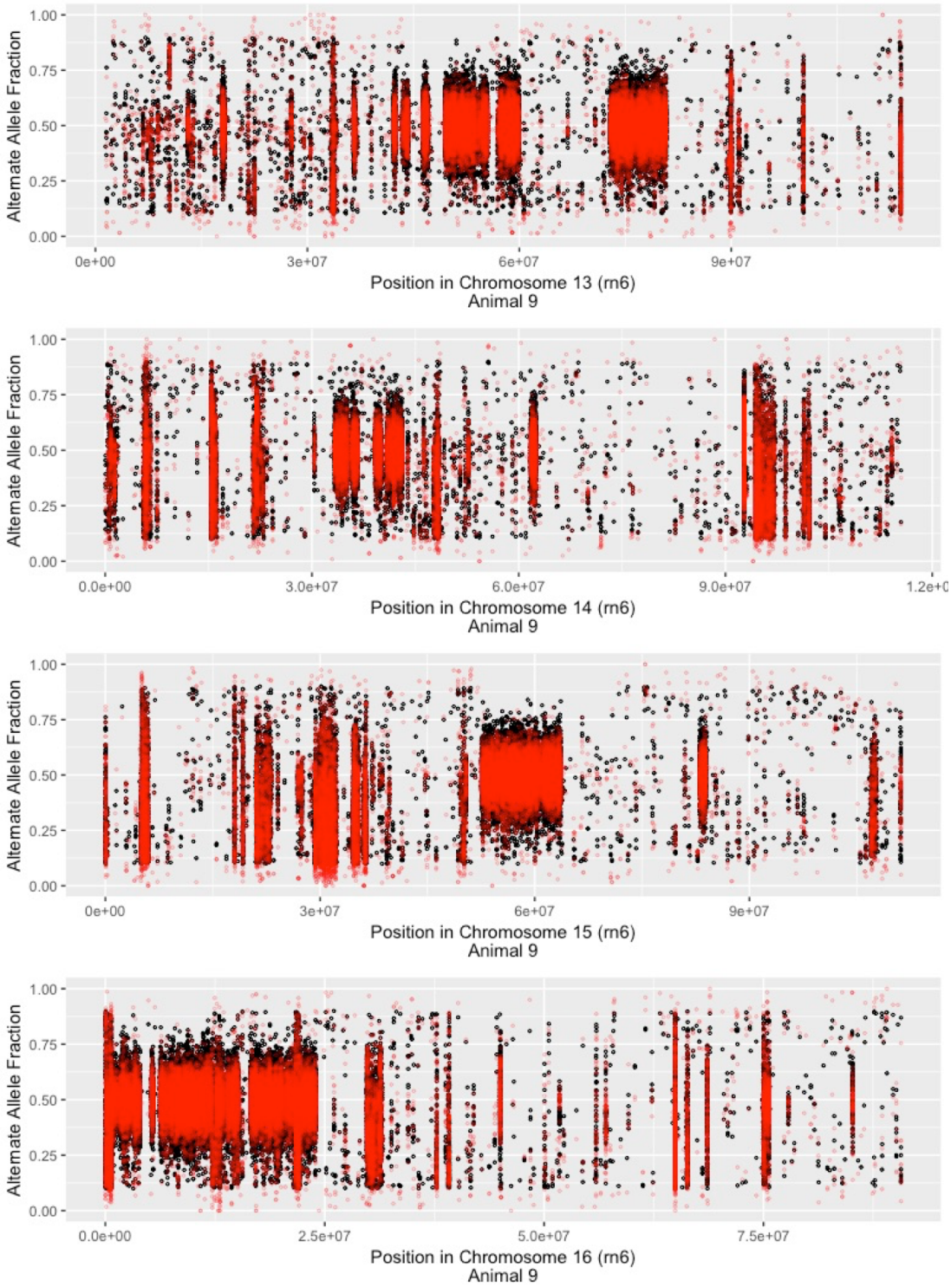

Figure 5.4.4 Screening for somatic LOH in a colonic tumour of the Pirc rat from chromosome 13 to 16.
The X-axis shows the position in the chromosome, where the somatic LOH was observed. The Y-axis shows the alternate allele fraction of different SNVs. The alternate allele fraction of SNVs from germline DNA is depicted with black dots, and tumour DNA with red dots. The discontinuous SNVs density was due to the absence of heterozygous SNVs the loci.

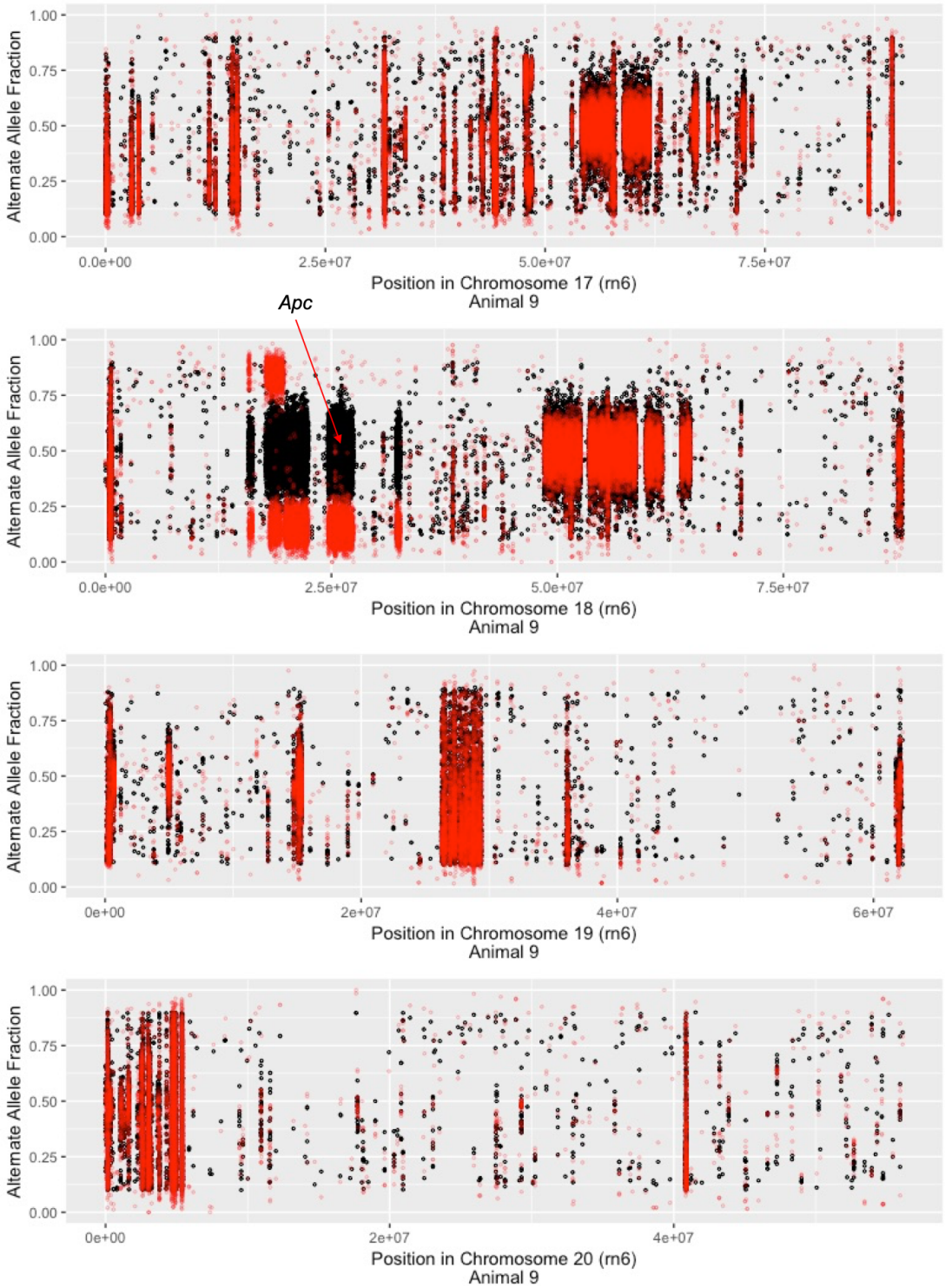

Figure 5.4.5 Screening for somatic LOH in a colonic tumour of the Pirc rat from chromosome 17 to 20.
The X-axis shows the position in chromosome 18, where the somatic LOH was observed. The Y-axis shows the alternate allele fraction of different SNVs. The alternate allele fraction of SNVs from germline DNA is depicted with black dots, and tumour DNA with red dots. The discontinuous SNVs density was due to the absence of heterozygous SNVs the loci. The location of the Apc gene in the genome is shown by a red arrow,

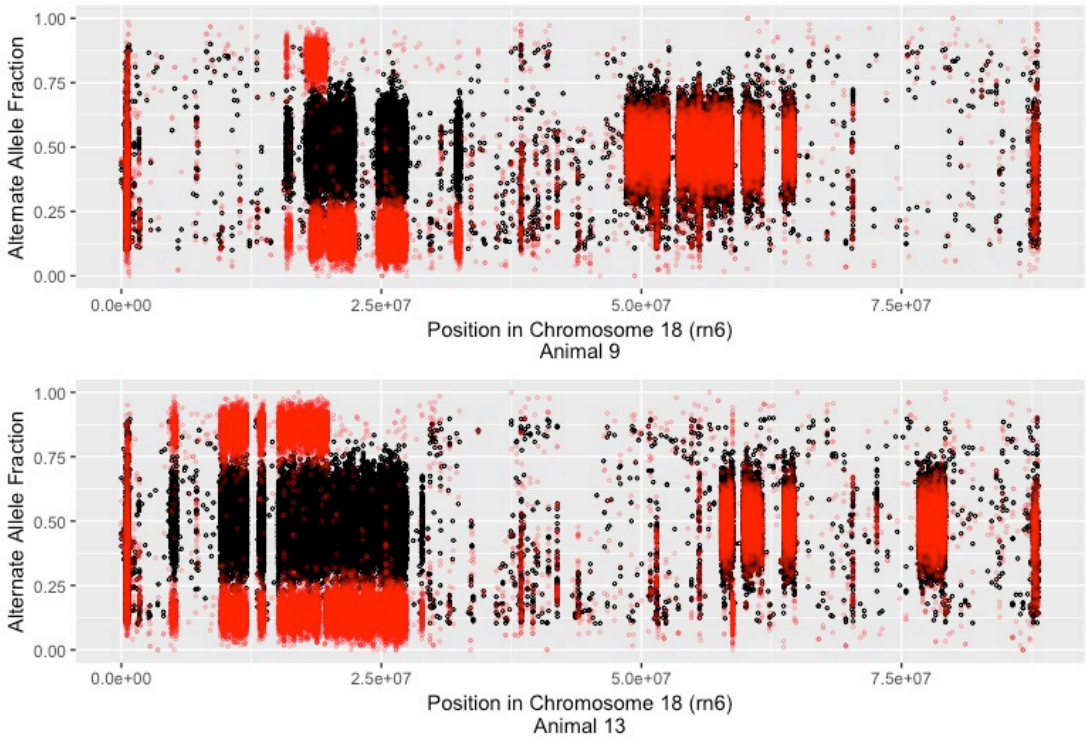

Figure 5.4.6 Somatic LOH in chromosome 18 in two colonic tumours of the Pirc rat. The X-axis shows the position in chromosome 18, where the somatic LOH was observed. The Y-axis shows the alternate allele fraction of different SNVs. The alternate allele fraction of SNVs from germline DNA is depicted with black dots, and tumour DNA with red dots. The discontinuous SNVs density was due to the absence of heterozygous SNVs the loci.

Table 5.4.1 Allelic and total sequencing depth at the Pirc locus

Sample	Allelic Depth at chr18:27100213		Total Depth at chr 18:27100213	Average Genome Coverage	Genotype
	A	T			
13PIRCADENOCARCINOMA	17	77	94	79.2	Heterozygous variant
13PIRCLIVER	21	20	41	37.28	Heterozygous variant
9PIRCADENOCARCINOMA	11	77	88	82.91	Heterozygous variant
9PIRCLIVER	23	16	39	44.18	Heterozygous variant
12WTLIVER	30	0	30	39.12	Homozygous variant

Circular binary segmentation analysis was performed using pseudo-odds to identify the start and end genomic coordinate of LOH based on the available SNVs (**Table 5.4.2**).

Table 5.4.2 Genomic coordinates of the somatic LOH regions in two Pirc rat colonic tumours.

Animal	Chromosome	Start	End	Number of supporting SNVs with skewed allele balance
9	18	15729175	19941051	2749
	18	19941529	21709559	2449
	18	21709847	28536304	4966
	18	32169550	32664300	576
13	18	4295334	18271596	9209
	18	18271848	18471422	322
	18	18471759	18781003	435
	18	18782934	18789866	13
	18	18791651	19559908	1151
	18	19560906	24209914	6542
	18	24210077	29216472	3865
	18	58729422	58789676	121

A ddPCR assay was designed to quantify allele balance of the single nucleotide variants in the Pirc locus and was then validated on gDNA extracted from the Pirc rat colonic tissues (**Figure 5.4.7**). A clear separation between negative and positive droplets can be observed at 55.8°C.

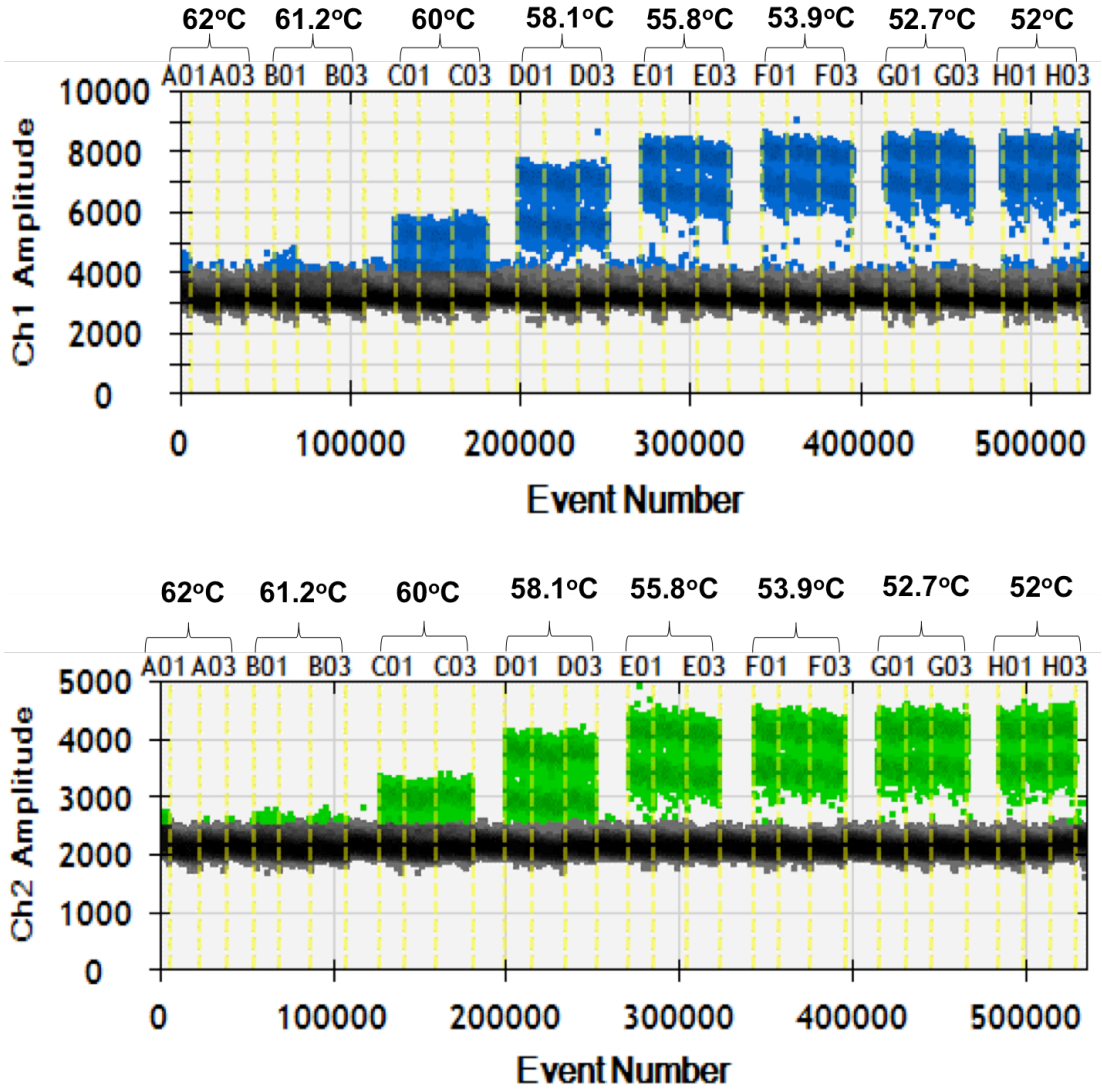

Figure 5.4.7 Validation of ddPCR assay to measure allele balance in the Pirc locus over a temperature gradient.
The X-axis showed number of droplets for each temperature replicate (event number), which was roughly less than 20000 droplets. The Y-axis showed the fluorescence amplitude emitted by each droplet.

Somatic LOH in the Pirc locus of two whole-genome sequenced tumours was validated using the custom-designed ddPCR assay. This assay was then performed on 46 colonic tumour samples from 21 different Pirc rats (**Figure 5.4.8**). All 46 tumours showed a shift in allele balance from 0.50 compared to their paired-germline DNA, indicating LOH in the Pirc locus of every tumour. The allele balance of the SNV in the tumours varied between 0.13 and 0.44.

This variation can most likely be explained by different tumour cellularity and tumour heterogeneity.

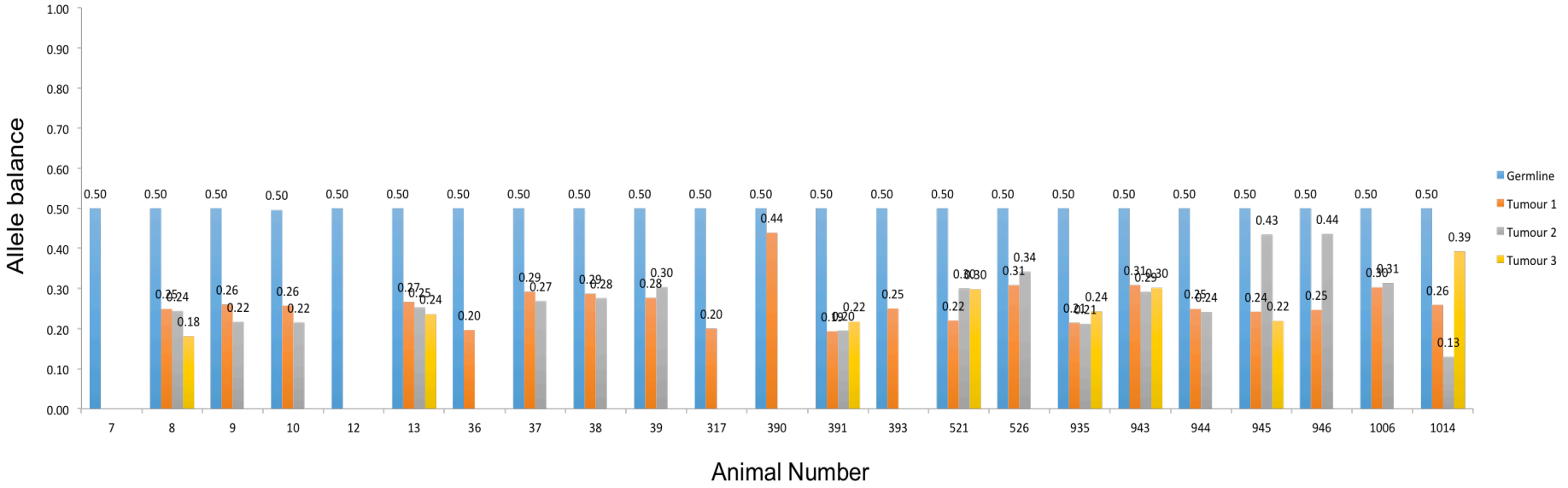

Figure 5.4.8 Screening of LOH across 46 colonic tumours from 21 ACI-Pirc rats and 2 ACI-WT rats.

Following successful LOH validation and screening in the Pirc locus of 46 tumours, 5 more ddPCR assays were designed to screen for LOH across chromosome 18. These ddPCR assays cover the beginning and end of the *Apc* gene, 2Mbps upstream and downstream of the Apc Pirc locus, as well as close to the long arm of chromosome 18 (**Table 5.4.3**). Across 46 colonic tumours, a majority of tumours (92.68%) have LOH spanning the whole *Apc* gene. The percentage of tumours with LOH dropped to 53.85% 2Mbps upstream of the Apc Pirc locus. No LOH was observed at the end of chromosome 18. These results suggest that heterozygous variants across the *Apc* gene of colonic tumours of the Pirc rat may be potential candidates of biomarkers to be detected in cfDNA of this model. Further, the difference of start and endpoint of LOH between tumours of the Pirc rat indicated the presence of inter-tumour heterogeneity in this model, despite all of the tumours arising from colon tissue.

Table 5.4.3 Screening of LOH across chromosome 18 of the Pirc rat cancer genome

Genomic Coordinate	Info	Percentage of Tumour with LOH
chr18:27100213	Pirc Locus	100.00
chr18:27043184	Beginning of Apc gene	92.68
chr18:27112121	End of Apc gene	92.68
chr18:28062387	28mb with LOH	92.68
chr18:25461411	25mb	53.85
chr18:87928418	End of chromosome	0.00

5.5 Screening for somatic single nucleotide, small insertion and deletion variants in the cancer genome of tumours from two Pirc rats

Screening for sSNVs was performed to identify single base substitutions in the Pirc rat cancer genome, which can ultimately be detected in cfDNA. Three somatic variant callers were used to identify sSNVs from the WGS alignment files. A total of 9010 and 9939 sSNVs (**Figure 5.5.1**) were identified in two colonic tumours of Pirc rat 9 and 13 respectively. Among these sSNVs, 469 and 1051 were found by at least two out of three of the somatic variant callers. These high confidence sSNVs were then functionally annotated based on putative impacts using snpEff, then compared to cancer gene list from the cancer gene consensus project of the COSMIC database to identify any sSNVs that had been reported in human cancer. Following functional annotation, the majority of the sSNVs had modifier impact, meaning that the change occurred at the non-coding region of the gene. Eleven variants (with modifier impact) occurred on the same gene from both samples. Of most interest were two missense variants found in tumour sample 9, which occurred in *Bcor* and *Plag1*, two previously reported cancer genes based on the COSMIC database.

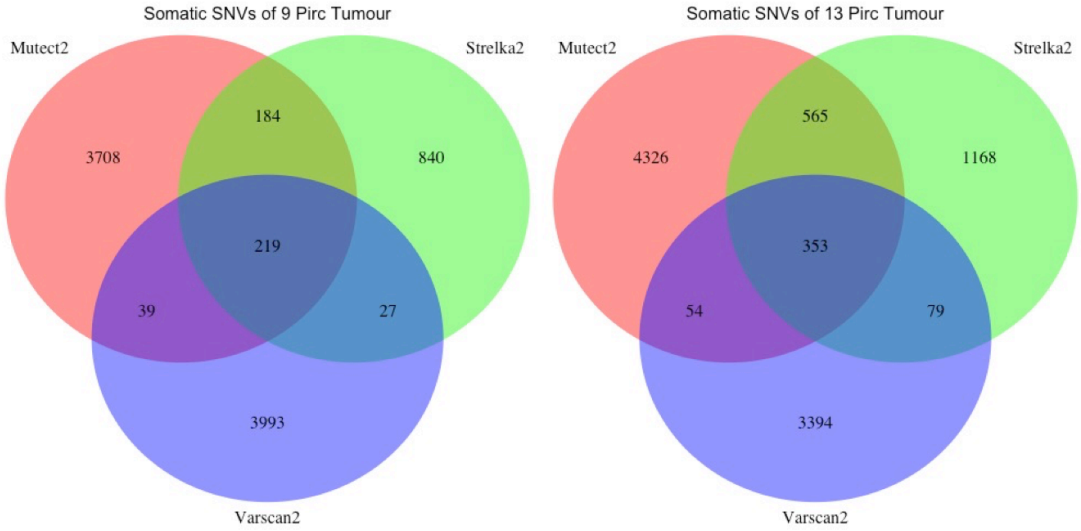

Figure 5.5.1 Somatic single nucleotide variants found by three somatic variant callers from whole genome sequence of the tumours of Pirc rats 9 and 13.

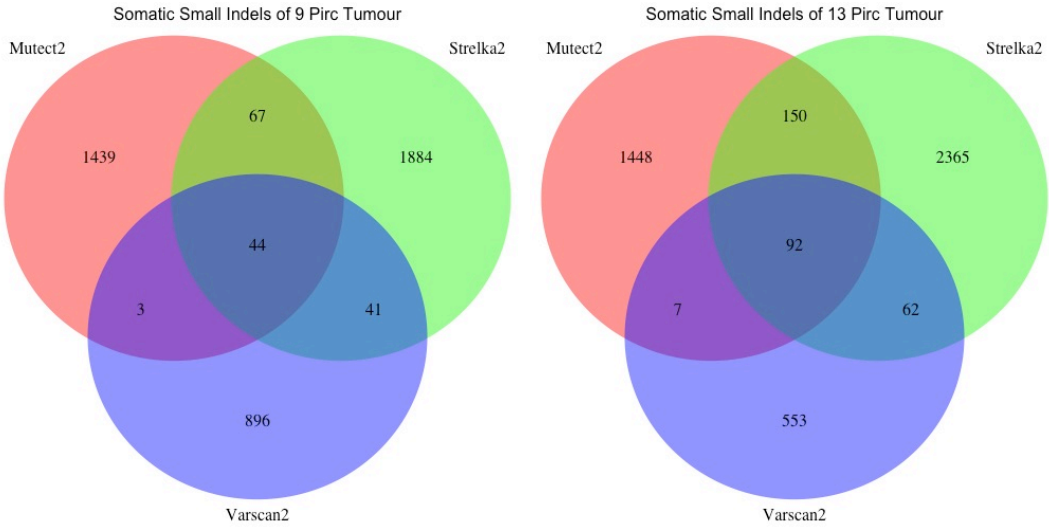

Figure 5.5.2 Somatic small insertion and deletion variants found by three somatic variant callers from whole genome sequence of the tumours of Pirc rats 9 and 13.

Screening for small insertion and deletion variants was performed to identify small insertions or deletions (2-50bp) in the Pirc rat cancer genome, which could ultimately be detected in cfDNA. Just like sSNVs, small indels were detected by using three somatic variant callers. A total of 4374 and 4827 small indels (**Figure 5.5.2**) were identified in two colonic tumours of Pirc rat 9 and 13 respectively. Among these small indels, 155 and 311 were found by at least

two out of three of the somatic variant callers. These high confidence small indels were then functionally annotated based on putative impacts using snpEff. Forty-eight small indel variants (with modifier impact) occurred on the same gene from both samples. Of most interest was a variant found in tumour sample 9, which occurred in Clp1.

5.6 Screening for structural and copy number variants in the Pirc rat cancer genome

Screening for structural and copy number variants were performed to identify structural deletions, duplications, inversions and other complex structural rearrangements that occurred in the Pirc rat cancer genome. A total of 601 and 581 structural variants were detected for sample 9 and 13 respectively (**Table 5.6.1**), a majority of which were intergenic variants. A list of genes that resides within genomic coordinates of exonic variants were obtained and compared to previously-known cancer genes from COSMIC database for human cancer. Of most interests were CCNE1, CEBPA, and MYOD1 genes. These genes were found to be duplicated in the tumour samples analysed. Additionally, copy number gains and overexpression of these genes were reported in the COSMIC database (human samples).

Table 5.6.1 Structural and copy number variants of the Pirc rat cancer genome.

Sample	Structural Variant Type	Variant Sub-type	Numbers
9	Duplication	Downstream	1
		Exonic	19
		Intergenic	19
		Intronic	2
		Upstream	1
	Deletion	Downstream	0
		Exonic	11
		Intergenic	45
		Intronic	3
		Upstream	1
	Inversion	Downstream	0
		Exonic	0
		Intergenic	1
		Intronic	1
		Upstream	0
	Other Complex Rearrangement	Downstream	4
		Exonic	2
		Intergenic	421
		Intronic	69
		Upstream	1
13	Duplication	Downstream	0
		Exonic	16
		Intergenic	26
		Intronic	2
		Upstream	0
	Deletion	Downstream	0
		Exonic	6
		Intergenic	43
		Intronic	4
		Upstream	0
	Inversion	Downstream	0
		Exonic	3
		Intergenic	1
		Intronic	0
		Upstream	0
	Other Complex Rearrangement	Downstream	1
		Exonic	2
		Intergenic	404
		Intronic	68
		Upstream	1
		Splicing	2
		UTR3	2

5.7 Discussion

In this chapter, WGS was performed to show somatic mutations in the cancer genome of the Pirc rat. Understanding somatic mutations in cancer genome may help understand how tumours develop, identify targets for therapy and novel biomarkers for detection of cancer. Specific for FAP patients, novel information on somatic mutations may help in diagnosing patients without genetic diagnosis to date. Further, genomic alterations found in this study serves as candidate biomarkers to be tested in cfDNA of the Pirc rat. Whole genome analysis of colonic tumours of the Pirc rat showed LOH in a large region of chromosome 18, including the *Apc* gene that may play a central role in tumourigenesis in the Pirc rat. Additionally, single nucleotide variants, small indels and structural variants were found in cancer genes linked with CRC in human.

Prior to whole genome analysis, the identity of colon tissue samples was validated using histological analysis of tissue sections stained with H&E. Histological analysis to date is still the golden standard for clinical diagnosis of tumours. This involves visual examination of cellular shapes to determine whether a tissue is cancerous [116]. In this project, the presence of irregularly-shaped intestinal glands confirmed the identity of colonic tumour tissues. This ensures the validity of subsequent analysis and helps to avoid false results from analysing wrong tissues.

WGS of tumour tissues in this study was performed at 60x average depth, as opposed to 30x in the germline. This deeper depth was selected because characterisation of somatic alterations from tumours can be challenging

because of variable tumour cellularity, polyploidy, and the presence of clonal heterogeneity. Sequencing depth of tumours at 30x is insufficient to detect variants with allele fraction of less than 15%, and at 50x of less than 10%. A sequencing depth of 200-300x is known to be required in order to facilitate variant discovery with less than 10% variant allele fraction [117].

WGS was performed on DNA from colonic tumours and matched germline DNA from liver in two individual rats (number 9 and 13). A matching germline DNA is considered essential for cancer genome sequencing to remove germline variants [117]. In this study, germline DNA was selected from liver and not adjacent normal colon tissue because adjacent normal colon tissues are predisposed to develop as tumours and may genetically have already accumulate mutations.

Sequence files generated in this study were of high quality, with base quality phred score above 40 across the sequencing reads. A phred score is a widely accepted measure of the probability that the base from sequencing is called correctly [118, 119]. Phred score of 40 in this study means that the probability of incorrect base call in this study is 1 in 10000 bases, supporting that variants in found in this study may be genuine variants and not due to sequencing errors.

Successful screening of LOH using WGS in this study extends LOH that has previously been found in the Pirc rat [115] at a single base resolution. Screening across different regions of chromosome 18 of the Pirc rat cancer genome with ddPCR confirmed that the *Apc* gene contained heterozygous variants that are promising biomarkers that could be detected in cfDNA.

Detection of these variants in cfDNA may extend the utility of the Pirc rat, not only as a model to study pathology of FAP, but also to study non-invasive detection of the disease. Variants of LOH may be potential candidates for non-invasive detection of cancer. This has previously been shown in patients with prostate cancer by the analysis of short tandem repeat sequences [120]. A careful design of assay to detect LOH in cfDNA is important, considering cfDNA of cancer patients consist of contributions from healthy tissues, such as hematopoietic and liver as previously found in Chapter 3, in addition to fragments originating from tumour cells.

During the screening of LOH using ddPCR at the Apc Pirc locus, a variability of alternate allele fraction in the tumours was observed, ranging from 0.18 to 0.44. This variability may be caused by differences in tumour cellularity, the fraction of malignant cells to healthy cells in tumour tissues analysed. Additionally, this variability may also be contributed from intra-tumour heterogeneity, the heterogeneity that arises in tumour cells within each tumour tissue.

Screening of LOH using ddPCR at several different loci of the chromosome 18 across 46 tumours also highlighted the presence of heterogeneity between colonic tumours of the Pirc rat (inter-tumour heterogeneity). LOH was found in all 46 colonic tumours at the Apc Pirc locus. The percentage of tumours with LOH across the whole *Apc* gene dropped to 92%. Less than 60% of the tumours have LOH at 2Mbps upstream of the Pirc locus.

The presence of inter-tumour heterozygosity would affect the detection LOH using cfDNA. The circulation of a tumour patient contains cfDNA arising from

normal cellular turnover and ctDNA from tumour tissues [22]. The detection of LOH in cfDNA would rely in the detection of relatively small shifts in allelic balance from 0.5 (heterozygous). In this case, the utilization of LOH biomarkers that are consistent between tumours would be beneficial compared to ones that are not. LOH biomarkers that are consistent between tumours would give stronger signal when detected in cfDNA compared to the inconsistent ones. It is, therefore, suggested to start the investigation of LOH biomarker utility in the Pirc rat using an assay designed for the Pirc locus.

In addition to confirmation of LOH in chromosome 18, analysis of single nucleotide variants, small indels, and structural variants were performed. Confident variants from these analyses were obtained by using multiple variant callers and intersecting the results to reduce false positive. The use of multiple variant callers to date is still considered an optimal strategy to improve analysis performance [117].

Following screening for somatic variants, the confident list of variants obtained in this study was compared with a list of variants from cancer gene census (COSMIC databse) in human [92, 93]. The list of cancer genes from cancer gene census consisted of expert-curated genes that drives human cancer and used as standard cancer genetics across basic researches. Tier 1 genes from the cancer gene census consisted of genes with documented activities relevant to cancer along with evidence of mutations in cancer that change the activity of the gene product that promotes tumourigenesis. Tier 2 genes consisted of genes with strong indications of roles in cancer with less extensive evidence available [92]. In this study, variants were found in genes included in

the cancer gene census. These genes are *Bcor*, *Plag1*, *Clp1*, *Ccne1*, *Cebpa*, and *Myod1*. Bcor is a tier 1 gene from cancer gene census. Inactivation of Bcor have been identified in various haematological cancers [121]. Further, single nucleotide variants of *Bcor* have been observed in colonic adenocarcinoma in cancer patients [122]. Overexpression of *Plag1* has been observed in various tumour types, such as adenoma of salivary glands, hepatoblastoma, lipoblastoma and acute myeloid leukaemia [123]. Mutation of the *Clp1* gene has been observed in human CRC [124] and in mouse model of CRC based on mutation in the *Apc* gene [125]. Copy number alterations of *Ccne1* is related to poor survival of bladder, breast and ovarian cancer patients [126]. Loss of function mutations of *Cebpa* contributes to the development of acute myeloid leukaemia [127]. *Myod1* is a putative tumour suppressor gene known to be disrupted in breast and CRC [124, 128].

Being a preliminary study, this study is mainly limited by the number of biological replicates for WGS, due to the limitation of cost. Somatic variants may be missed because of the lack of replicates. Additional samples for WGS would be required depending on the aim of follow up studies, such as to discover novel genes that is involved in tumourigenesis of the Pirc rat. Further, time limitation prevented the study of variants found in this project as biomarkers to be used in cfDNA.

Finally, to develop the Pirc rat as a model organism to study non-invasive detection of cancer, several steps need to be taken. First, cfDNA from the Pirc rat needs to be extracted and quantified. Following cfDNA extraction and quantification, an in-silico calculation may be performed to estimate the

detection threshold of LOH ddPCR assays and should take into account the sensitivity of LOH ddPCR assays. A proof-of-concept experiment can be performed to analyse LOH at the Pirc locus on a few cfDNA samples to demonstrate the possibility of showing LOH in cfDNA using LOH ddPCR assay. Upon successful demonstration, this experiment can then be extended to many animals at different loci with LOH across chromosome 18.

Chapter 6 General Discussion

This PhD project sets out to perform genomic analysis in mouse models to investigate the tissue origins of cfDNA and in a rat model to screen for potential candidate biomarkers of liquid biopsy to detect CRC. Various findings were discovered throughout the study. This includes utilisation of ddPCR assay to measure Cre-loxP recombination in tissue-specific reporter mice, the first absolute and most direct measurement of the tissue origins of cfDNA in healthy mice and in a representative pathological condition of APAP overdose, the potential of cfDNA analyses as a biomarker of tissue injury in APAP overdose patients, and the first CRC genome sequencing of the Pirc rat. In this chapter, these key findings that form the novelty of this PhD thesis, their significance, as well as their immediate and longer-term implications in the field of medicine were discussed.

6.1 Validation of ddPCR assay to quantify Cre-loxP recombination in tissue-specific reporter mice

To perform absolute measurement of the tissue origins of mouse cfDNA, Cre-loxP recombination in cfDNA of tissue-specific reporter mice was quantified. An assay to quantify Cre-loxP recombination in gDNA and cfDNA was designed and validated using ddPCR. This assay was adapted from a qPCR approach [82]. Successful utilisation of ddPCR to quantify Cre-loxP recombination is a substantial improvement from other methods, such as southern blotting and qPCR. Quantitation of Cre-loxP recombination in cells and tissues is important because it affects the accuracy of an experiment [82].

In this study, for example, quantitation of Cre-recombination in cardiomyocyte-specific reporter mice showed that there is background recombination coming from skeletal muscle cells (less than 5%), but not from other cells (less than 1%). Compared to qPCR, quantitation of Cre-loxP recombination with ddPCR is reference free. This reduces human error, for example during generation of standard curve, and allows for more accurate quantitation. Further, the input DNA requirement for ddPCR is low, unlike southern blotting that requires more than 5ug of DNA [82]. Based on these advantages, it is recommended to utilise ddPCR to quantify Cre-loxP recombination when required.

6.2 Absolute measurement of the tissue origins of cfDNA in healthy mice

Using tissue-specific reporter mice and a ddPCR assay to quantify Cre-loxP recombination, an absolute and most direct measurement the tissue origins of cfDNA was carried out. This includes the contribution from six cell types: myeloid, lymphoid, erythroid cardiomyocyte and striated muscle cells. Myeloid, lymphoid and erythroid cells were major contributors of cfDNA origin in the healthy state, contributing up to ~79% of the cfDNA pool in healthy mice aged 10-12 week old. A minor contribution of ~4% was observed from hepatocytes. No contribution was detected from muscle cells. These results complement and are concordant with previous findings in human using relative measurement method, such as comparison of nucleosomal mapping and tissue-specific methylation patterns [14, 24–26]. Around 17% of tissue contribution in this study was unknown, and likely to originate from endothelial cells, or other cell types that have yet to be investigated. In the short term,

these findings are useful for academics and laboratories in designing cfDNA assays to detect cancer based on DNA methylation or nucleosomal mappings. An understanding of the composition of cfDNA in the healthy state is useful to be included in a bioinformatics pipeline for target discovery, aiming to reduce false positive results originating from non-target tissues. Aside from cancer, this knowledge may also be useful for detection of other pathological conditions, where there is any deviation from the rate of cellular turnover in the healthy state. In the longer term, these findings may contribute toward successful design and implementation of robust, non-invasive assays based on cfDNA for various conditions.

6.3 Measurement of the tissue origins of cfDNA following tissue injury

Following the absolute measurement of the tissue origins of cfDNA in healthy mice, an absolute measurement of the tissue origins of cfDNA was also performed in hepatocyte-specific reporter mice following APAP overdose. This experiment was the first to demonstrate the utility of tissue-specific reporter mice in measuring the tissue origins of cfDNA in a pathological condition. Following APAP overdose, an increase of the contribution of hepatocytes from ~4% to 77% was successfully detected. This demonstration opens up the possibility of using tissue-specific reporter mouse models for measurement of the tissue origins of cfDNA in other pathological conditions.

6.4 cfDNA analysis as a biomarker of tissue injury in APAP overdose patients

During the investigation of the tissue origins of cfDNA following APAP overdose in mice, it was found that the total concentration cfDNA by more than 100-fold and contribution of cfDNA from hepatocytes increased by ~20-fold. This raised a question whether the analysis of cfDNA can be used as a biomarker of tissue injury in APAP overdose patients. To answer this question, a preliminary experiment was performed to investigate the ability of both the total concentration of cfDNA and liver-derived cfDNA to distinguish APAP overdose patients and healthy volunteers. The total concentration of cfDNA robustly distinguished APAP overdose patients from healthy volunteers, irrespective of an apparent liver injury based on protein liver function tests, such as ALT and AST. This indicates that the total concentration of cfDNA may be a more sensitive biomarker of tissue injury in APAP overdose patients. Unlike the total concentration of cfDNA, liver-derived cfDNA can only distinguish patients with apparent liver injury from healthy volunteers, and not patients without apparent liver injury from healthy volunteers. This may indicate that liver-derived cfDNA reflect a more specific injury in the liver as opposed to effects of APAP overdose in other systems in the body. When compared with other biomarkers of liver function tests, the total concentration of cfDNA is highly concordant with a biomarker of mitochondrial damage, GLDH, whereas liver derived cfDNA showed concordant results with ALT. Altogether, this preliminary study is the first to show the potential of cfDNA analyses in detecting APAP overdose patients. Investigation of alternative biomarkers of APAP overdose is important because current APAP overdose

markers, such as serum ALT and APAP concentration, lack sensitivity and specificity when measured soon after overdose such as at initial presentation to hospital [102]. In the short term, these findings may serve as a basis for academics and clinicians in designing bigger studies to investigate the utility of cfDNA analyses as a biomarker of tissue injury in APAP overdose patients. In the long term, this study may contribute toward a novel assay to detect APAP overdose in the clinic.

6.5 Genomic landscape of CRC in the Pirc rat

To screen for potential candidate biomarkers for the detection of CRC, WGS was performed in two paired colonic tumours and normal colon of the Pirc rat. Whole genome sequence analysis showed LOH in chromosome 18 of the Pirc rat cancer genome. Validation of LOH across chromosome 18 in 46 colonic tumours using ddPCR showed that LOH occurred for all tumour samples in the Apc Pirc locus (chr18:27100213) and in the majority (~97%) of the tumours across the whole *Apc* gene. This indicated that LOH in the Apc gene may be an important mechanism in tumourigenesis in the Pirc rat and showed that SNVs in the Apc gene of colonic tumours of the Pirc rat may be promising candidates for detection of CRC. Aside from LOH, various somatic alterations, including sSNVs, small indels, and structural as well as copy number variants were analysed. Of most interest were notable variants found in both this study and previously reported in the COSMIC cancer gene census database [92], including sSNVs in *Bcor* and *Plag1*, a small indel in *Clp1*, and structural variants in *Ccne1*, *Cebpa* and *Myod1*. These findings, although preliminary,

are the first to show the landscape of cancer genome in the Pirc rat. Further, this study provides the first characterisation of LOH in the Pirc rat at a single base resolution across the genome. These findings formed a basis for follow up investigations in the Pirc rat that aims to expand the utility of the Pirc rat to as a platform to study liquid biopsy, and as a model to potentially discover novel genes for detection of CRC, specifically FAP, in human. Discovery of novel genes for FAP is important because a proportion of FAP patients are negative for mutations in the *APC* gene, and do not have a genetic diagnosis [112]. In the longer term, these findings may contribute to a novel non-invasive assay for detection of CRC in the clinic.

6.6 Follow up studies

Taking into considerations of these key findings, several follow-up investigations can be performed. A follow-up absolute measurement of tissue origins of cfDNA in the healthy state may be performed using a different experimental design, such as utilisation of tissue-specific knock-in mice, instead of deletion of a reporter gene, utilisation of Cre recombinase driven by other promoters to investigate different cell types, or utilisation of different floxed gene. These studies may show the presence or absence of contributions from other tissues to expand understanding of the tissue origins of cfDNA, and further demonstrate the robustness of mouse models for studies of cfDNA.

A follow up study to investigate the tissue origins of cfDNA may be performed using tissue-specific reporter mice in various physiological conditions, such as

ageing, following exercise and physical stress, or pathological conditions, such as following myocardial infarction, alcohol poisoning, or other tissue injuries. These studies may serve as an exploration of cfDNA analyses as a biomarker for these conditions. Specifically for APAP overdose, a follow-up investigation in mice may be beneficial when performed using different doses of administration and at earlier as well as later timepoints following APAP overdose. This study may show the dynamics of different APAP overdose biomarkers, including cfDNA analyses and liver function tests, in response to APAP dosage at different timepoint.

Following preliminary demonstration of cfDNA analyses as a biomarker of tissue injury in APAP overdose patients, a follow-up study with more samples may be performed to show clinical validity and utility of cfDNA analyses. This study may also show the specificity and sensitivity of the total concentration of cfDNA and liver-derived cfDNA as a biomarker of APAP overdose. To allow for successful implementation of cfDNA analyses as a biomarker of APAP overdose in the clinic, establishing a point-of-care assay based on cfDNA may be beneficial. This assay would need to have a relatively quick turnaround time from sample collection to results interpretation by clinicians, preferably comparable to current liver function tests such as ALT and AST. At the moment, measurement of cfDNA concentration can be performed in a relatively short time using fluorometry, such as Qubit dsDNA assay used in quantitation of gDNA in this study. An alternative to Qubit, quantitation of cfDNA using qPCR is substantially slower but provides better specificity because it is PCR based.

An ideal point-of-care tests based on cfDNA would need to have both specificity and quick turnaround time.

Following preliminary analysis of the colonic tumours of the Pirc rat, two studies can be performed. First, more colonic tumour samples may be sequenced to allow for robust discovery of novel genes to study carcinogenesis of FAP. Identification of novel genes in rat model may benefit in controlling external variables and in a scenario where paired germline genome and sample collection is restricted. The knowledge obtained from successful identification of novel genes in FAP may be translated to help diagnose APC mutation negative patients [112]. Sequencing additional samples may also provide more candidate markers of liquid biopsy to be detected in cfDNA to demonstrate the utility of the Pirc rat as a model of CRC liquid biopsy.

Other future work that is not directly related to key findings in this study may also be performed. Upon the realisation of low input amount of cfDNA from mouse and from healthy individuals, as well as heterogeneity of tumour samples and low mutant allele fractions of variants, studies of tumour allele enrichment may enhance the progress of liquid biopsy.

References

1. Heitzer E, Haque IS, Roberts CES, Speicher MR. Current and future perspectives of liquid biopsies in genomics-driven oncology. Nat Rev Genet. 2019;20:71–88.

2. Marrugo-Ramírez J, Mir M, Samitier J. Blood-Based Cancer Biomarkers in Liquid Biopsy: A Promising Non-Invasive Alternative to Tissue Biopsy. Int J Mol Sci. 2018;19:2877. doi:10.3390/ijms19102877.

3. Jung A, Kirchner T. Liquid Biopsy in Tumor Genetic Diagnosis. Dtsch Arztebl Int. 2018;115:169–74. doi:10.3238/arztebl.2018.0169.

4. De Rubis G, Rajeev Krishnan S, Bebawy M. Liquid Biopsies in Cancer Diagnosis, Monitoring, and Prognosis. Trends Pharmacol Sci. 2019.

5. Fettke H, Kwan EM, Azad AA. Cell-free DNA in cancer: current insights. Cell Oncol. 2019;42:13–28. doi:10.1007/s13402-018-0413-5.

6. Goossens N, Nakagawa S, Sun X, Hoshida Y. Cancer biomarker discovery and validation. Transl Cancer Res. 2015;4:256–69. doi:10.3978/j.issn.2218-676X.2015.06.04.

7. Schwarzenbach H, Hoon DSB, Pantel K. Cell-free nucleic acids as biomarkers in cancer patients. Nat Rev Cancer. 2011;11:426–37.

8. Levenson V V. DNA methylation as a universal biomarker. Expert Rev Mol Diagn. 2010;10:481–8. doi:10.1586/erm.10.17.

9. Griffiths A, Miller J, Suzuki D, Lewontin R, Gelbart W. An Introduction to Genetic Analysis. 7th edition. New York: W.H. Freeman; 2000.

10. Lodish H, Berk A, Zipursky S, Matsudaira P, Baltimore D, Darnell J. Molecular Cell Biology. 4th editio. New York: W.H. Freeman; 2000. https://www.ncbi.nlm.nih.gov/books/NBK21475/.

11. Pfeifer GP. Defining Driver DNA Methylation Changes in Human Cancer. Int J Mol Sci. 2018;19:1166. doi:10.3390/ijms19041166.

12. Lokk K, Modhukur V, Rajashekar B, Märtens K, Mägi R, Kolde R, et al. DNA methylome profiling of human tissues identifies global and tissue-specific methylation patterns. Genome Biol. 2014;15:r54–r54. doi:10.1186/gb-2014-15-4-r54.

13. Diaz LAJ, Bardelli A. Liquid biopsies: genotyping circulating tumor DNA. J Clin Oncol Off J Am Soc Clin Oncol. 2014;32:579–86.

14. Snyder MW, Kircher M, Hill AJ, Daza RM, Shendure J. Cell-free DNA comprises an in vivo nucleosome footprint that informs its tissues-of-origin. Cell. 2016;164:57–68.

15. Kang S, Li Q, Chen Q, Zhou Y, Park S, Lee G, et al. CancerLocator: non-invasive cancer diagnosis and tissue-of-origin prediction using methylation profiles of cell-free DNA. Genome Biol. 2017;18:53. doi:10.1186/s13059-017-1191-5.

16. Mandel P, Métais P. Les acides nucléiques du plasma sanguin chez l'homme. C R Seances Soc Biol Fil. 1948;142:241–3.

17. Martignano F. Cell-free DNA: an overview of sample types and isolation procedures. Methods Mol Biol. 2019;1909:13–27.

18. Thierry AR, El Messaoudi S, Gahan PB, Anker P, Stroun M. Origins, structures, and functions of circulating DNA in oncology. Cancer Metastasis Rev. 2016;35:347–76.

19. Aucamp J, Bronkhorst AJ, Badenhorst CPS, Pretorius PJ. The diverse origins of circulating cell-free DNA in the human body: a critical re-evaluation of the literature. Biol Rev Camb Philos Soc. 2018;93:1649–83.

20. Serpas L, Chan RWY, Jiang P, Ni M, Sun K, Rashidfarrokhi A, et al. Dnase1l3 deletion causes aberrations in length and end-motif frequencies in plasma DNA. Proc Natl Acad Sci. 2019;116:641 LP – 649. doi:10.1073/pnas.1815031116.

21. Fan HC, Blumenfeld YJ, Chitkara U, Hudgins L, Quake SR. Analysis of the size distributions of fetal and maternal cell-free DNA by paired-end sequencing. Clin Chem. 2010;56:1279–86.

22. Crowley E, Di Nicolantonio F, Loupakis F, Bardelli A. Liquid biopsy: monitoring cancer-genetics in the blood. Nat Rev Clin Oncol. 2013;10:472. https://doi.org/10.1038/nrclinonc.2013.110.

23. Lui YYN, Chik K-W, Chiu RWK, Ho C-Y, Lam CWK, Lo YMD. Predominant hematopoietic origin of cell-free DNA in plasma and serum after sex-mismatched bone marrow transplantation. Clin Chem. 2002;48:421–7.

24. Lam WKJ, Gai W, Sun K, Wong RSM, Chan RWY, Jiang P, et al. DNA of erythroid origin is present in human plasma and informs the types of anemia. Clin Chem. 2017;63:1614–23.

25. Sun K, Jiang P, Chan KCA, Wong J, Cheng YKY, Liang RHS, et al. Plasma DNA tissue mapping by genome-wide methylation sequencing for noninvasive prenatal, cancer, and transplantation assessments. Proc Natl Acad Sci U S A. 2015;112:E5503-12.

26. Moss J, Magenheim J, Neiman D, Zemmour H, Loyfer N, Korach A, et al. Comprehensive human cell-type methylation atlas reveals origins of circulating cell-free DNA in health and disease. Nat Commun. 2018;9:5068.

27. Lo YM, Corbetta N, Chamberlain PF, Rai V, Sargent IL, Redman CW, et al. Presence of fetal DNA in maternal plasma and serum. Lancet. 1997;350:485–7.

28. Stroun M, Anker P, Maurice P, Lyautey J, Lederrey C, Beljanski M. Neoplastic Characteristics of the DNA Found in the Plasma of Cancer Patients. Oncology. 1989;46:318–22. doi:10.1159/000226740.

29. Vermeesch JR, Voet T, Devriendt K. Prenatal and pre-implantation genetic diagnosis. Nat Rev Genet. 2016;17:643. https://doi.org/10.1038/nrg.2016.97.

30. Renga B. Non invasive prenatal diagnosis of fetal aneuploidy using cell free fetal DNA. Eur J Obstet Gynecol Reprod Biol. 2018;225:5–8.

31. Knight SR, Thorne A, Lo Faro ML. Donor-specific Cell-free DNA as a Biomarker in Solid Organ Transplantation. A Systematic Review. Transplantation. 2019;103:273–83.

32. Lehmann-Werman R, Neiman D, Zemmour H, Moss J, Magenheim J, Vaknin-Dembinsky A, et al. Identification of tissue-specific cell death using methylation patterns of circulating DNA. Proc Natl Acad Sci U S A. 2016;113:E1826-34.

33. Zemmour H, Planer D, Magenheim J, Moss J, Neiman D, Gilon D, et al. Non-invasive detection of human cardiomyocyte death using methylation patterns of circulating DNA. Nat Commun. 2018;9:1443. doi:10.1038/s41467-018-03961-y.

34. Neumann MHD, Bender S, Krahn T, Schlange T. ctDNA and CTCs in Liquid Biopsy - Current Status and Where We Need to Progress. Comput Struct Biotechnol J. 2018;16:190–5. doi:10.1016/j.csbj.2018.05.002.

35. Markus H, Contente-Cuomo T, Farooq M, Liang WS, Borad MJ, Sivakumar S, et al. Evaluation of pre-analytical factors affecting plasma DNA analysis. Sci Rep. 2018;8:7375. doi:10.1038/s41598-018-25810-0.

36. Grölz D, Hauch S, Schlumpberger M, Guenther K, Voss T, Sprenger-Haussels M, et al. Liquid Biopsy Preservation Solutions for Standardized Pre-Analytical Workflows-Venous Whole Blood and Plasma. Curr Pathobiol Rep. 2018;6:275–86. doi:10.1007/s40139-018-0180-z.

37. Zinkova A, Brynychova I, Svacina A, Jirkovska M, Korabecna M. Cell-free DNA from human plasma and serum differs in content of telomeric sequences and its ability to promote immune response. Sci Rep. 2017;7:2591. doi:10.1038/s41598-017-02905-8.

38. Trigg RM, Martinson LJ, Parpart-Li S, Shaw JA. Factors that influence quality and yield of circulating-free DNA: A systematic review of the methodology literature. Heliyon. 2018;4:e00699.

39. Sherwood JL, Corcoran C, Brown H, Sharpe AD, Musilova M, Kohlmann A. Optimised Pre-Analytical Methods Improve KRAS Mutation Detection in Circulating Tumour DNA (ctDNA) from Patients with Non-Small Cell Lung Cancer (NSCLC). PLoS One. 2016;11:e0150197.

40. El Messaoudi S, Rolet F, Mouliere F, Thierry AR. Circulating cell free DNA: Preanalytical considerations. Clin Chim Acta. 2013;424:222–30.

41. Fong SL, Zhang JT, Lim CK, Eu KW, Liu Y. Comparison of 7 Methods for Extracting Cell-Free DNA from Serum Samples of Colorectal Cancer Patients. Clin Chem. 2009;55:587 LP – 589. doi:10.1373/clinchem.2008.110122.

42. Sorber L, Zwaenepoel K, Deschoolmeester V, Roeyen G, Lardon F, Rolfo C, et al. A Comparison of Cell-Free DNA Isolation Kits: Isolation and Quantification of Cell-Free DNA in Plasma. J Mol Diagn. 2017;19:162–8.

43. Heitzer E, Ulz P, Geigl JB. Circulating Tumor DNA as a Liquid Biopsy for Cancer. Clin Chem. 2015;61:112 LP – 123. doi:10.1373/clinchem.2014.222679.

44. Devonshire AS, Whale AS, Gutteridge A, Jones G, Cowen S, Foy CA, et al. Towards standardisation of cell-free DNA measurement in plasma: controls for extraction efficiency, fragment size bias and quantification. Anal Bioanal Chem. 2014;406:6499–512. doi:10.1007/s00216-014-7835-3.

45. Thierry AR, Mouliere F, Gongora C, Ollier J, Robert B, Ychou M, et al. Origin and quantification of circulating DNA in mice with human colorectal cancer xenografts. Nucleic Acids Res. 2010;38:6159–75.

46. Umetani N, Giuliano AE, Hiramatsu SH, Amersi F, Nakagawa T, Martino S, et al. Prediction of Breast Tumor Progression by Integrity of Free Circulating DNA in Serum. J Clin Oncol. 2006;24:4270–6. doi:10.1200/JCO.2006.05.9493.

47. Lapin M, Oltedal S, Tjensvoll K, Buhl T, Smaaland R, Garresori H, et al. Fragment size and level of cell-free DNA provide prognostic information in patients with advanced pancreatic cancer. J Transl Med. 2018;16:300. doi:10.1186/s12967-018-1677-2.

48. Newman AM, Bratman S V, To J, Wynne JF, Eclov NCW, Modlin LA, et al. An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage. Nat Med. 2014;20:548–54. doi:10.1038/nm.3519.

49. Forshew T, Murtaza M, Parkinson C, Gale D, Tsui DWY, Kaper F, et al. Noninvasive identification and monitoring of cancer mutations by targeted deep sequencing of plasma DNA. Sci Transl Med. 2012;4:136ra68.

50. Little S. Amplification-refractory mutation system (ARMS) analysis of point mutations. Curr Protoc Hum Genet. 2001;Chapter 9:Unit 9.8.

51. Takai E, Yachida S. Circulating tumor DNA as a liquid biopsy target for detection of pancreatic cancer. World J Gastroenterol. 2016;22:8480–8. doi:10.3748/wjg.v22.i38.8480.

52. Tong Y, Shen S, Jiang H, Chen Z. Application of Digital PCR in Detecting Human Diseases Associated Gene Mutation. Cell Physiol Biochem. 2017;43:1718–30. doi:10.1159/000484035.

53. Song C, Liu Y, Fontana R, Makrigiorgos A, Mamon H, Kulke MH, et al. Elimination of unaltered DNA in mixed clinical samples via nuclease-assisted minor-allele enrichment. Nucleic Acids Res. 2016;44:e146–e146. doi:10.1093/nar/gkw650.

54. Liu Y, Song C, Ladas I, Fitarelli-Kiehl M, Makrigiorgos GM. Methylation-sensitive enrichment of minor DNA alleles using a double-strand DNA-specific nuclease. Nucleic Acids Res. 2016;45:e39–e39. doi:10.1093/nar/gkw1166.

55. Chan LL, Jiang P. Bioinformatics analysis of circulating cell-free DNA sequencing data. Clin Biochem. 2015;48:962–75.

56. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr Protoc Bioinforma. 2013;43:11.10.1-33.

57. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat Biotechnol. 2013;31:213. https://doi.org/10.1038/nbt.2514.

58. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25:1754–60.

59. Nakagawa H, Fujita M. Whole genome sequencing analysis for cancer genomics and precision medicine. Cancer Sci. 2018;109:513–22. doi:10.1111/cas.13505.

60. Krueger F, Andrews SR. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics. 2011;27:1571–2.

61. Hansen KD, Langmead B, Irizarry RA. BSmooth: from whole genome bisulfite sequencing reads to differentially methylated regions. Genome Biol. 2012;13:R83. doi:10.1186/gb-2012-13-10-r83.

62. Li Y, Tollefsbol TO. DNA methylation detection: bisulfite genomic sequencing analysis. Methods Mol Biol. 2011;791:11–21. doi:10.1007/978-1-61779-316-5_2.

63. Kurdyukov S, Bullock M. DNA Methylation Analysis: Choosing the Right Method. Biology (Basel). 2016;5:3. doi:10.3390/biology5010003.

64. Sauer B, Henderson N. Site-specific DNA recombination in mammalian cells by the Cre recombinase of bacteriophage P1. Proc Natl Acad Sci U S A. 1988;85:5166–70.

65. Vandamme TF. Use of rodents as models of human diseases. J Pharm Bioallied Sci. 2014;6:2–9. doi:10.4103/0975-7406.124301.

66. Kim H, Kim M, Im S-K, Fang S. Mouse Cre-LoxP system: general principles to determine tissue-specific roles of target genes. Lab Anim Res. 2018;34:147–59. doi:10.5625/lar.2018.34.4.147.

67. Yue F, Cheng Y, Breschi A, Vierstra J, Wu W, Ryba T, et al. A comparative encyclopedia of DNA elements in the mouse genome. Nature. 2014;515:355–64. doi:10.1038/nature13992.

68. Chinwalla AT, Cook LL, Delehaunty KD, Fewell GA, Fulton LA, Fulton RS, et al. Initial sequencing and comparative analysis of the mouse genome. Nature. 2002;420:520–62. doi:10.1038/nature01262.

69. Irving AA, Yoshimi K, Hart ML, Parker T, Clipson L, Ford MR, et al. The utility of Apc-mutant rats in modeling human colon cancer. Dis Model Mech. 2014;7:1215–25. doi:10.1242/dmm.016980.

70. Iannaccone PM, Jacob HJ. Rats! Dis Model Mech. 2009;2:206–10. doi:10.1242/dmm.002733.

71. Gasparello J, Allegretti M, Tremante E, Fabbri E, Amoreo CA, Romania P, et al. Liquid biopsy in mice bearing colorectal carcinoma xenografts: gateways regulating the levels of circulating tumor DNA (ctDNA) and miRNA (ctmiRNA). J Exp Clin Cancer Res. 2018;37:124. doi:10.1186/s13046-018-0788-1.

72. Muhanna N, Di Grappa MA, Chan HHL, Khan T, Jin CS, Zheng Y, et al. Cell-Free DNA Kinetics in a Pre-Clinical Model of Head and Neck Cancer. Sci Rep. 2017;7:16723. doi:10.1038/s41598-017-17079-6.

73. Clausen BE, Burkhardt C, Reith W, Renkawitz R, Forster I. Conditional gene targeting in macrophages and granulocytes using LysMcre mice. Transgenic Res. 1999;8:265–77.

74. de Boer J, Williams A, Skavdis G, Harker N, Coles M, Tolaini M, et al. Transgenic mice with hematopoietic and lymphoid specific expression of Cre. Eur J Immunol. 2003;33:314–25.

75. Jiao K, Kulessa H, Tompkins K, Zhou Y, Batts L, Baldwin HS, et al. An essential role of Bmp4 in the atrioventricular septation of the mouse heart. Genes Dev. 2003;17:2362–7.

76. Postic C, Shiota M, Niswender KD, Jetton TL, Chen Y, Moates JM, et al. Dual roles for glucokinase in glucose homeostasis as determined by liver and pancreatic beta cell-specific gene knock-outs using Cre recombinase. J Biol Chem. 1999;274:305–15.

77. Bruning JC, Michael MD, Winnay JN, Hayashi T, Horsch D, Accili D, et al. A muscle-specific insulin receptor knockout exhibits features of the metabolic syndrome of NIDDM without altering glucose tolerance. Mol Cell. 1998;2:559–69.

78. Heinrich AC, Pelanda R, Klingmuller U. A mouse model for visualization and conditional mutations in the erythroid lineage. Blood. 2004;104:659–66.

79. Muzumdar MD, Tasic B, Miyamichi K, Li L, Luo L. A global double-fluorescent Cre reporter mouse. Genesis. 2007;45:593–605.

80. Bateman DN, Dear JW, Thanacoody HKR, Thomas SHL, Eddleston M, Sandilands EA, et al. Reduction of adverse effects from intravenous acetylcysteine treatment for paracetamol poisoning: a randomised controlled trial. Lancet. 2014;383:697–704.

81. Li P, Ning J, Luo X, Du H, Zhang Q, Zhou G, et al. New method to preserve the original proportion and integrity of urinary cell-free DNA. J Clin Lab Anal. 2018;:e22668.

82. Weis B, Schmidt J, Lyko F, Linhart HG. Analysis of conditional gene deletion using probe based Real-Time PCR. BMC Biotechnol. 2010;10:75.

83. Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, et al. Primer3--new capabilities and interfaces. Nucleic Acids Res. 2012;40:e115.

84. Gai W, Ji L, Lam WKJ, Sun K, Jiang P, Chan AWH, et al. Liver- and colon-specific DNA methylation markers in plasma for investigation of colorectal cancers with or without liver metastases. Clin Chem. 2018;64:1239–49.

85. Atanur SS, Diaz AG, Maratou K, Sarkis A, Rotival M, Game L, et al. Genome sequencing reveals loci under artificial selection that underlie disease phenotypes in the laboratory rat. Cell. 2013;154:691–703.

86. Gibbs RA, Weinstock GM, Metzker ML, Muzny DM, Sodergren EJ, Scherer S, et al. Genome sequence of the Brown Norway rat yields insights into mammalian evolution. Nature. 2004;428:493–521.

87. Havlak P, Chen R, Durbin KJ, Egan A, Ren Y, Song X-Z, et al. The Atlas genome assembly system. Genome Res. 2004;14:721–32.

88. Olshen AB, Bengtsson H, Neuvial P, Spellman PT, Olshen RA, Seshan VE. Parent-specific copy number in paired tumor-normal studies using circular binary segmentation. Bioinformatics. 2011;27:2038–46.

89. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res. 2012;22:568–76.

90. Kim S, Scheffler K, Halpern AL, Bekritsky MA, Noh E, Källberg M, et al. Strelka2: fast and accurate calling of germline and somatic variants. Nat Methods. 2018;15:591–4. doi:10.1038/s41592-018-0051-x.

91. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin). 2012;6:80–92. doi:10.4161/fly.19695.

92. Sondka Z, Bamford S, Cole CG, Ward SA, Dunham I, Forbes SA. The COSMIC Cancer Gene Census: describing genetic dysfunction across all human cancers. Nat Rev Cancer. 2018;18:696–705. doi:10.1038/s41568-018-0060-1.

93. Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, et al. COSMIC: the Catalogue Of Somatic Mutations In Cancer. Nucleic Acids Res. 2019;47:D941–7.

94. Layer RM, Chiang C, Quinlan AR, Hall IM. LUMPY: a probabilistic framework for structural variant discovery. Genome Biol. 2014;15:R84. doi:10.1186/gb-2014-15-6-r84.

95. Chen X, Schulz-Trieglaff O, Shaw R, Barnes B, Schlesinger F, Kallberg M, et al. Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. Bioinformatics. 2016;32:1220–2.

96. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38:e164–e164. doi:10.1093/nar/gkq603.

97. McLellan MA, Rosenthal NA, Pinto AR. Cre-loxP-Mediated Recombination: General Principles and Experimental Considerations. Curr Protoc Mouse Biol. 2017;7:1–12.

98. Mort RL, Ford MJ, Sakaue-Sawano A, Lindstrom NO, Casadio A, Douglas AT, et al. Fucci2a: a bicistronic cell cycle reporter that allows Cre mediated tissue specific expression in mice. Cell Cycle. 2014;13:2681–96. doi:10.4161/15384101.2015.945381.

99. Bayascas JR, Sakamoto K, Armit L, Arthur JSC, Alessi DR. Evaluation of approaches to generation of tissue-specific knock-in mice. J Biol Chem. 2006;281:28772–81.

100. Lehmann-Werman R, Magenheim J, Moss J, Neiman D, Abraham O, Piyanzin S, et al. Monitoring liver damage using hepatocyte-specific methylation markers in cell-free circulating DNA. JCI insight. 2018;3.

101. Beck J, Oellerich M, Schulz U, Schauerte V, Reinhard L, Fuchs U, et al. Donor-Derived Cell-Free DNA Is a Novel Universal Biomarker for Allograft Rejection in Solid Organ Transplantation. Transplant Proc. 2015;47:2400–3.

102. Dear JW, Clarke JI, Francis B, Allen L, Wraight J, Shen J, et al. Risk stratification after paracetamol overdose using mechanistic biomarkers: results from two prospective cohort studies. Lancet Gastroenterol Hepatol. 2018;3:104–13.

103. Athersuch TJ, Antoine DJ, Boobis AR, Coen M, Daly AK, Possamai L, et al. Paracetamol metabolism, hepatotoxicity, biomarkers and therapeutic interventions: a perspective. Toxicol Res (Camb). 2018;7:347–57. doi:10.1039/c7tx00340d.

104. Yan M, Huo Y, Yin S, Hu H. Mechanisms of acetaminophen-induced liver injury and its implications for therapeutic interventions. Redox Biol. 2018;17:274–83.

105. Mossanen JC, Tacke F. Acetaminophen-induced acute liver injury in mice. Lab Anim. 2015;49 1_suppl:30–6. doi:10.1177/0023677215570992.

106. Duan L, Davis JS, Woolbright BL, Du K, Cahkraborty M, Weemhoff J, et al. Differential susceptibility to acetaminophen-induced liver injury in sub-strains of C57BL/6 mice: 6N versus 6J. Food Chem Toxicol. 2016;98 Pt B:107–18. doi:10.1016/j.fct.2016.10.021.

107. Kennon-McGill S, McGill MR. Extrahepatic toxicity of acetaminophen: critical evaluation of the evidence and proposed mechanisms. J Clin Transl Res. 2018;3.

108. Sherwood K, Weimer ET. Characteristics, properties, and potential applications of circulating cell-free dna in clinical diagnostics: a focus on transplantation. J Immunol Methods. 2018;463:27–38.

109. Antoine DJ, Dear JW, Lewis PS, Platt V, Coyle J, Masson M, et al. Mechanistic biomarkers provide early and sensitive detection of acetaminophen-induced acute liver injury at first presentation to hospital. Hepatology. 2013;58:777–87.

110. McGill MR, Sharpe MR, Williams CD, Taha M, Curry SC, Jaeschke H. The mechanism underlying acetaminophen-induced hepatotoxicity in humans and mice involves mitochondrial damage and nuclear DNA fragmentation. J Clin Invest. 2012;122:1574–83.

111. Spanaki C, Kotzamani D, Petraki Z, Drakos E, Plaitakis A. Expression of human GLUD1 and GLUD2 glutamate dehydrogenases in steroid producing tissues. Mol Cell Endocrinol. 2015;415:1–11.

112. Talseth-Palmer BA. The genetic basis of colonic adenomatous polyposis syndromes. Hered Cancer Clin Pract. 2017;15:5.

113. Lamlum H, Papadopoulou A, Ilyas M, Rowan A, Gillet C, Hanby A, et al. APC mutations are sufficient for the growth of early colorectal adenomas. Proc Natl Acad Sci U S A. 2000;97:2225–8. doi:10.1073/pnas.040564697.

114. Amos-Landgraf JM, Kwong LN, Kendziorski CM, Reichelderfer M, Torrealba J, Weichert J, et al. A target-selected Apc-mutant rat kindred enhances the modeling of familial human colon cancer. Proc Natl Acad Sci U S A. 2007;104:4036–41. doi:10.1073/pnas.0611690104.

115. Amos-Landgraf JM, Irving AA, Hartman C, Hunter A, Laube B, Chen X, et al. Monoallelic silencing and haploinsufficiency in early murine intestinal neoplasms. Proc Natl Acad Sci U S A. 2012;109:2060–5. doi:10.1073/pnas.1120753109.

116. He L, Long LR, Antani S, Thoma GR. Histology image analysis for carcinoma detection and grading. Comput Methods Programs Biomed. 2012;107:538–56. doi:10.1016/j.cmpb.2011.12.007.

117. Griffith M, Miller CA, Griffith OL, Krysiak K, Skidmore ZL, Ramu A, et al. Optimizing cancer genome sequencing and analysis. Cell Syst. 2015;1:210–23. doi:10.1016/j.cels.2015.08.015.

118. Ewing B, Hillier L, Wendl MC, Green P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res. 1998;8:175–85.

119. Liao P, Satten GA, Hu Y-J. PhredEM: a phred-score-informed genotype-calling approach for next-generation sequencing studies. Genet Epidemiol. 2017;41:375–87. doi:10.1002/gepi.22048.

120. Seyedolmohadessin S-M, Akbari MT, Nourmohammadi Z, Basiri A, Pourmand G. Detection of Loss of Heterozygosity (LOH) Using Circulating Cell-free DNA (cfDNA) by Fluorescence-based Multiplex PCR for Identification of Patients With Prostate Cancer. Appl Immunohistochem Mol Morphol AIMM. 2018;26:749–59.

121. Tanaka T, Nakajima-Takagi Y, Aoyama K, Tara S, Oshima M, Saraya A, et al. Internal deletion of BCOR reveals a tumor suppressor function for BCOR in T lymphocyte malignancies. J Exp Med. 2017;214:2901 LP – 2913. doi:10.1084/jem.20170167.

122. Zehir A, Benayed R, Shah RH, Syed A, Middha S, Kim HR, et al. Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. Nat Med. 2017;23:703–13.

123. Akhtar M, Holmgren C, Göndör A, Vesterlund M, Kanduri C, Larsson C, et al. Cell type and context-specific function of PLAG1 for IGF2 P3 promoter activity. Int J Oncol. 2012;41:1959–66. doi:10.3892/ijo.2012.1641.

124. Sjoblom T, Jones S, Wood LD, Parsons DW, Lin J, Barber TD, et al. The consensus coding sequences of human breast and colorectal cancers. Science. 2006;314:268–74.

125. March HN, Rust AG, Wright NA, ten Hoeve J, de Ridder J, Eldridge M, et al. Insertional mutagenesis identifies multiple networks of cooperating genes driving intestinal tumorigenesis. Nat Genet. 2011;43:1202–9. doi:10.1038/ng.990.

126. Zhao H, Wang J, Zhang Y, Yuan M, Yang S, Li L, et al. Prognostic Values of CCNE1 Amplification and Overexpression in Cancer Patients: A Systematic Review and Meta-analysis. J Cancer. 2018;9:2397–407. doi:10.7150/jca.24179.

127. Lourenço AR, Coffer PJ. A tumor suppressor role for C/EBPα in solid tumors: more than fat and blood. Oncogene. 2017;36:5221. https://doi.org/10.1038/onc.2017.151.

128. Ahuja N, Li Q, Mohan AL, Baylin SB, Issa JP. Aging and DNA methylation in colorectal mucosa and cancer. Cancer Res. 1998;58:5489–94.

129. Limdi JK, Hyde GM. Evaluation of abnormal liver function tests. Postgrad Med J. 2003;79:307 LP – 312. doi:10.1136/pmj.79.932.307.

130. Schomaker S, Warner R, Bock J, Johnson K, Potter D, Van Winkle J, et al. Assessment of emerging biomarkers of liver injury in human subjects. Toxicol Sci. 2013;132:276–83.

Appendix

A. Supplementary Figures

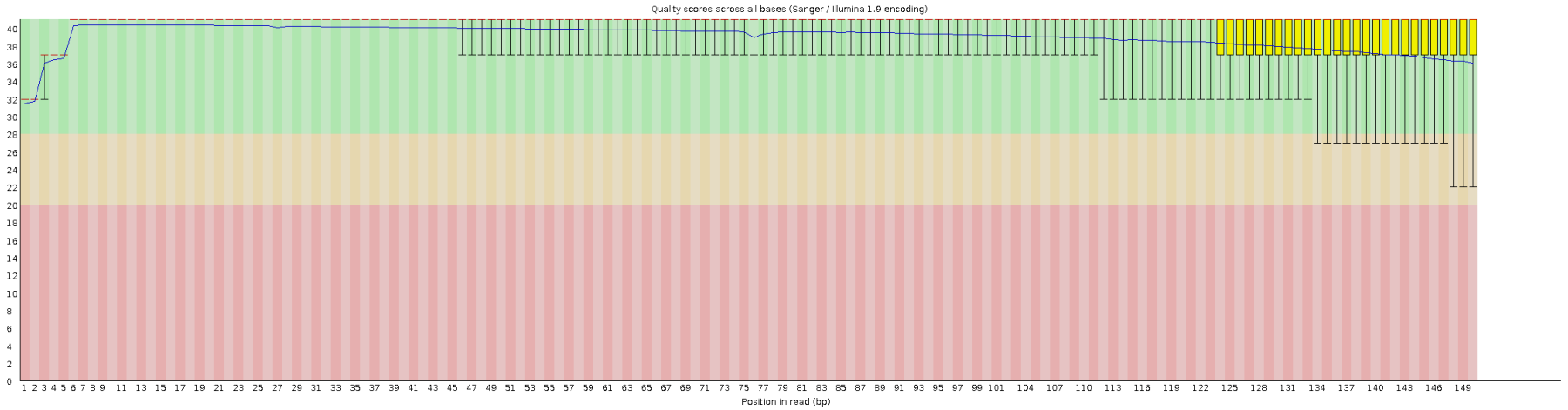

Figure A.1 A representative plot of base quality of sequences across sequencing read.
X axis showed position in read (bp) with maximum of 150bp. Y axis showed average Phred score. An average of more than 30 Phred score was observed across the sequencing read.

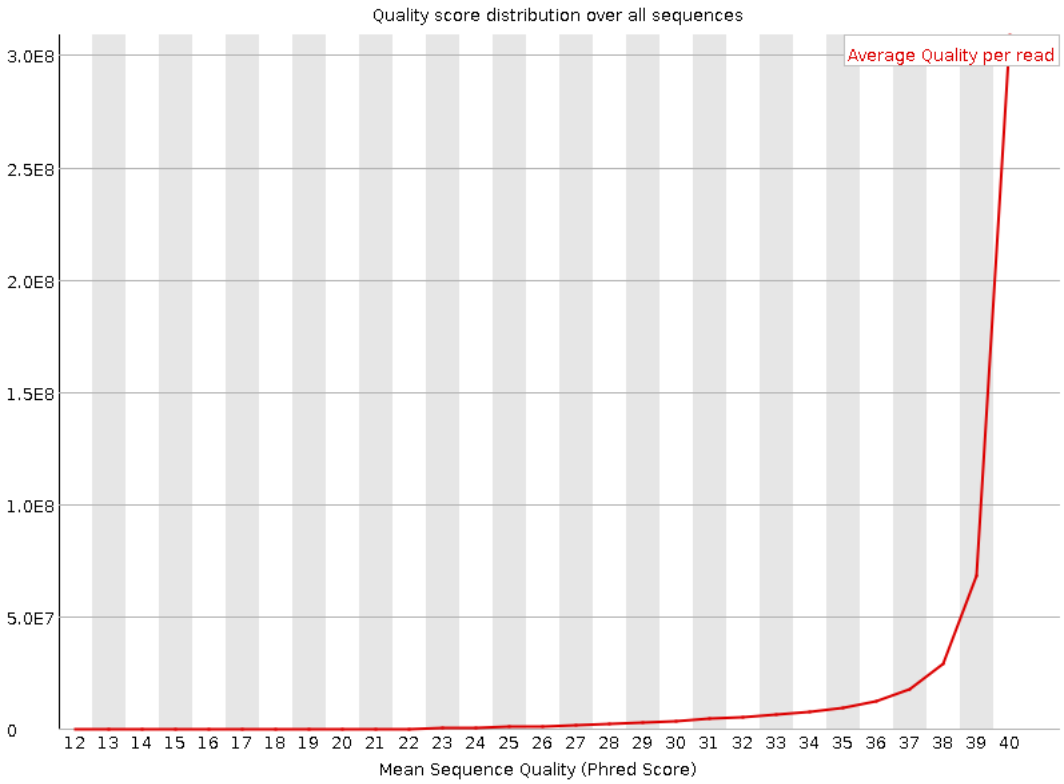

Figure A.2 A representative plot of distribution of Phred quality score over all sequences.
X axis showed mean sequence quality (Phred score), Y axis showed the number of sequences. Majority of sequences has quality score of more than 39.

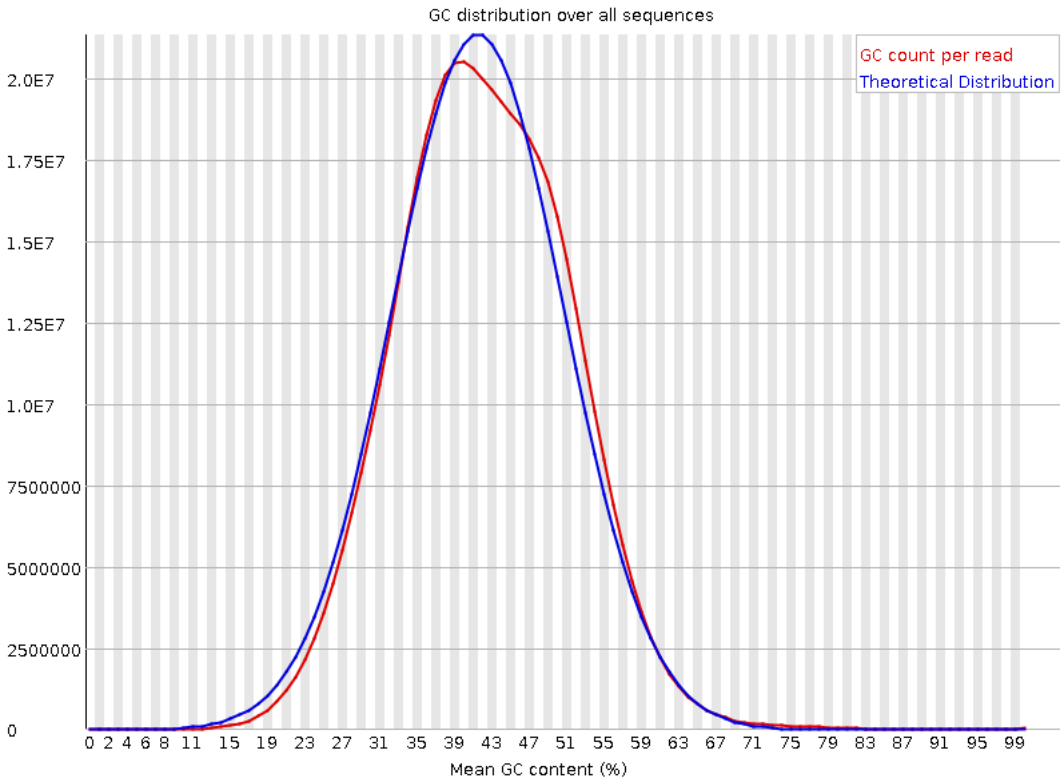

Figure A.3 A representative plot of GC content distribution from whole genome sequencing.
X axis showed mean GC content (%), Y axis showed the number of sequences. A bell curved distribution of GC content was observed (shown by red line), similar to theoretical GC content distribution (shown by blue line), indicating normal GC content distribution in the sequence data.

B. Supplementary Tables

Table B.1 Measurements of serum biomarkers from APAP overdose patients and healthy volunteers.

Sample	Cohort	Biomarkers							
		cfDNA (ng/ml)	Bilirubin (umol/L)	ALT (IU/l)	AST (IU/l)	ALP (IU/l)	ALB (g/l)	GLDH (IU/l)	Percentage of cfDNA from liver
JD1	APAP Overdose Patients	10882.5	22.3	3391	2483	145	40.8	653	95.5
JD2		255.7	<2.5	5019	1794	280	26.8	504	16.4
JD3		18998.8	18.7	2251	3049	150	31.7	1056	97.7
JD4		7351.0	11.9	2795	3114	291	28.9	583	80.7
JD5		192.9	7.2	11	36	91	38.4	11	0.1
JD6		1168.9	<2.5	17	26	117	36.4	18	0.3
JD7		846.4	107	25	85	127	43.4	68	1.2
JD8		787.5	7.2	17	26	96	39.3	33	0.0
PS1	Healthy Volunteers	188.2	22.6	12	29	117	44.5	11.7	0.0
PS2		20.5	5.3	14	29	77	37.2	9.1	0.0
PS3		11.7	4.3	12	52	81	38.7	9.1	1.5
PS4		49.7	7.8	21	43	<50	41.9	11.7	0.0
PS5		26.3	12.7	21	27	101.4	43.7	6.5	1.2
PS6		64.7	2.6	34	32	76	38.8	10	0.5
PS7		10.5	3.9	16	36	104	40.6	7.8	0.0
PS8		8.4	6.6	22	35	105	44.3	7.8	0.0
PS9		21.4	6.5	17	33	126	44.3	13	0.9
PS10		24.9	9.8	65	26	91	40.3	3.9	0.4
PS11		13.6	13.0	18	33	135	42.3	3.9	0.0

Normal range for serum biomarker [129, 130]: Bilirubin (2-17 μmol/l), ALT (0-45 IU/l), AST (0-35 IU/l), ALP (30-120 IU/l), ALB/Albumin (40-60 g/l), GLDH (1-10 IU/l).

C. Statistical analysis

C.1 Comparison of body weight from different tissue-specific reporter mouse lines

A Shapiro-Wilk test was performed to investigate data distribution on the animal body weight data set. The null hypothesis of Shapiro-Wilk test is the data follow a normal distribution. The alternate hypothesis is the data do not follow a normal distribution. Shapiro-Wilk test showed p value of 0.3 ($p > 0.05$). Therefore, null hypothesis was not rejected. The data follow normal distribution.

Despite an indication of normal distribution, a non-parametric Kruskall-Wallis test was selected as opposed to one-way analysis of variance (ANOVA) because to perform parametric test ANOVA, the data for each group should contain at least 15 data point for 2-9 groups. The null hypothesis Kruskall-Wallis test is all medians are equal. The alternate hypothesis is at least one median is different Analysis with Kruskall-Wallis test showed p value of 9.5×10^{-7} ($p < 0.05$). Therefore, null hypothesis was rejected. At least one median in the dataset was different.

A Mann-Whitney U test was selected to find the group with different median (**Table C.1.1**) in the dataset as opposed to student t test due to sample size being less than 15 (**Table C.1.2**). The null hypothesis of Mann-Whitney U test is the median between two groups are equal. The alternate hypothesis is the median between two groups are not equal. Grouping of tissue-specific reporter lines were performed based on p-value from Mann-Whitney U test (**Table C.1.2**). Group A consisted of Alb Cre, EpoR Cre, and MCK Cre. Group

B consisted of EpoR Cre, MCK Cre, hCD2-iCre, and no Cre. Group C consisted of no Cre, LysM Cre, and cTnT Cre.

Table C.1.1 Median and standard deviation of the body weight of tissue-specific reporter mice.

Mouse line	Number of mice	Animal body weight (g)	
		Median	Standard deviation
Alb Cre	13	27.0	2.4
cTnT Cre	12	23.3	1.7
EpoR Cre	20	26.2	2.1
hCD2-iCre	12	25.5	1.4
LysM Cre	16	23.0	2.0
MCK Cre	10	25.4	1.2
no Cre	10	23.7	1.7

Table C.1.2 p-values from comparison of mouse body weight between reporter lines with a Mann-Whitney U test.

p-value	Alb Cre	cTnT Cre	Epor Cre	hCD2-iCre	LysM Cre	MCK Cre	no Cre
Alb Cre	NA	0.00	0.44	0.02	0.00	0.05	0.00
cTnT Cre	0.00	NA	0.00	0.01	0.61	0.01	0.16
Epor Cre	0.44	0.00	NA	0.06	0.00	0.06	0.00
hCD2-iCre	0.02	0.01	0.06	NA	0.01	1.00	0.17
LysM Cre	0.00	0.61	0.00	0.01	NA	0.02	0.11
MCK Cre	0.05	0.01	0.06	1.00	0.02	NA	0.04
no Cre	0.00	0.16	0.00	0.17	0.11	0.04	NA

NA: not available. p-values in yellow are less than 0.05.

C.2 Confirmation of Cre recombination in mouse tissues

Data distribution of percentage recombination (1lox%) in mouse tissues was analysed using Shapiro-Wilk test (**Table C.2.1**). The null hypothesis of Shapiro-Wilk test is the data follow a normal distribution. The alternate hypothesis is the data do not follow a normal distribution. Majority of datasets

followed normal distribution, except for ear notch of cTnT Cre line, ear notch of Alb Cre line, whole blood of LysM Cre line, and ear notch of MCK Cre line. Comparison using parametric student t test requires both datasets used in comparison to be normally distributed and with at least 15 replicates for each dataset. Considering these statistical assumptions, a non-parametric Mann-Whitney U test was selected to compare tissues within each reporter lines.

Table C.2.1 Summary of percentage recombination and data distribution based on Shapiro-Wilk test in mouse tissues.

Reporter Line	Tissue	n	1lox%		Shapiro-Wilk p-value
			Median	Standard deviation	
cTnT Cre	heart	12	39.0	3.8	0.65
	ear notch		0.8	0.7	0.03
EpoR Cre	bone marrow	20	28.1	7.1	0.81
	ear notch		1.6	0.7	0.06
Alb Cre	liver	13	71.6	6.7	0.09
	ear notch		0.0	0.3	0.00
hCD2-iCre	whole blood	12	71.5	5.7	0.90
	ear notch		3.0	0.7	0.44
LysM Cre	whole blood	16	14.1	4.8	0.01
	ear notch		4.3	1.7	0.20
MCK Cre	skeletal muscle	10	25.9	7.9	0.82
	ear notch		0.2	0.2	0.02

p-values in yellow are less than 0.05.

Comparison between tissues within each reporter line showed a significant difference (**Table C.2.2**). This confirmed the presence of recombination of the floxed *mT/mG* gene in expected tissues for every mouse reporter line (p-value < 0.05).

Table C.2.2 p-value from a Mann-Whitney U test between tissues with and without recombination.

Reporter line	Mann-Whitney U p-value
cTnT Cre	7.4×10^{-7}
EpoR Cre	6.7×10^{-8}
Alb Cre	1.5×10^{-5}
hCD2-iCre	3.6×10^{-5}
LysM Cre	3.3×10^{-9}
MCK Cre	0.01

p-values in yellow are less than 0.05.

C.3 Analysis of cfDNA and other liver biomarkers in clinical samples

C.3.1 Comparison of total cfDNA in clinical samples

Data distribution of total cfDNA measured by qPCR of the *ACTB* gene was analysed using a Shapiro-Wilk test. The null hypothesis of Shapiro-Wilk test is the data follow a normal distribution. The alternate hypothesis is the data do not follow a normal distribution. Shapiro-Wilk test showed p-value of 5.459×10^{-7} (p-value < 0.05). Therefore, null hypothesis was rejected. The total cfDNA data do not follow normal distribution.

A non-parametric Mann-Whitney U test was used to analysed total cfDNA between healthy volunteers (HV), APAP overdose patients with normal ALT (AON), and with high ALT (AOH) (**Table C.3.1.1**). A significant difference in median value of total cfDNA was observed between HV and AON, and HV and AOH, but not between AON and AOH.

Table C.3.1.1 p-value from comparison of total cfDNA between clinical samples.

p-value	HV	AON	AOH
HV	NA	0.001	0.001
AON	0.001	NA	0.2
AOH	0.001	0.2	NA

HV: healthy volunteers; AON: APAP overdose patients with normal ALT; AOH: APAP overdose patients with high ALT. p-value shown in yellow is less than 0.05.

C.3.2 Comparison of liver-specific cfDNA in clinical samples

Data distribution of liver-specific cfDNA in clinical samples was analysed using a Shapiro-Wilk test. The null hypothesis of Shapiro-Wilk test is the data follow a normal distribution. The alternate hypothesis is the data do not follow a normal distribution. Shapiro-Wilk test showed p-value of 5.459×10^{-7} (p-value < 0.05). Therefore, null hypothesis was rejected. The liver-specific cfDNA data do not follow normal distribution.

A non-parametric Mann-Whitney U test was used to analysed liver-specific cfDNA between healthy volunteers (HV), APAP overdose patients with normal ALT (AON), and with high ALT (AOH) (**Table C.3.2.1**). A significant difference in median value of liver-specific cfDNA was observed between AOH and AON, as well as AOH and HV, but not between AON and HV.

Table C.3.2.1 p-value from comparison of liver-specific cfDNA between clinical samples.

p-value	HV	AON	AOH
HV	NA	0.78	0.004
AON	0.78	NA	0.03
AOH	0.004	0.03	NA

HV: healthy volunteers; AON: APAP overdose patients with normal ALT; AOH: APAP overdose patients with high ALT. p-value shown in yellow is less than 0.05.

C.3.3 Associations of biomarkers in clinical samples

Associations between total cfDNA and protein liver function biomarkers, as well as between liver-specific cfDNA and other liver function biomarkers were performed to compare the performance of total cfDNA and liver-specific cfDNA with protein liver function biomarkers, namely ALT and GLDH.

A Spearman rho test was used for these association tests because both total cfDNA and liver-specific cfDNA do not follow normal data distribution (see section **C.3.1** and **C.3.2**). A strong positive association was observed between total cfDNA and GLDH (p-value < 0.01). Positive association was observed between liver-specific cfDNA and ALT, as well as liver-specific cfDNA and GLDH (p-value < 0.05) (**Table C.3.3.1**).

Table C.3.3.1 Spearman rho and p-value from analysis of associations between biomarkers in clinical samples.

cfDNA analysis	Liver protein biomarker	Spearman rho	p-value
total cfDNA	ALT	0.44	0.059
	GLDH	0.85	3.38×10^{-6}
liver-specific cfDNA	ALT	0.57	0.01
	GLDH	0.59	0.01

p-value shown in yellow is less than 0.05.